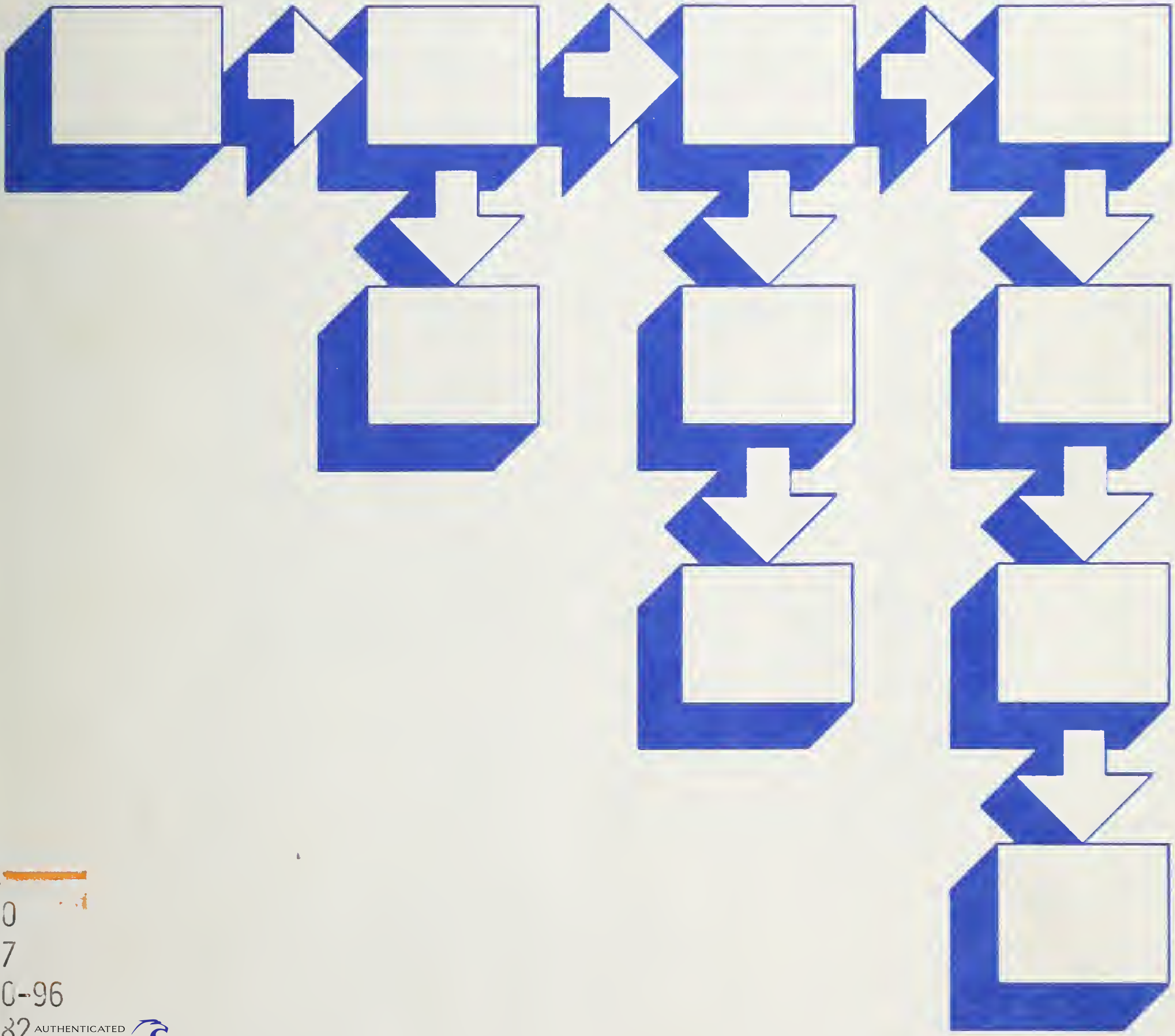
of Commerce

National Bureau
of Standards

Computer Science and Technology

NBS Special Publication 500-96

The Selection of Local Area Computer Networks



NATIONAL BUREAU OF STANDARDS

The National Bureau of Standards[1] was established by an act of Congress on March 3, 1901. The Bureau's overall goal is to strengthen and advance the Nation's science and technology and facilitate their effective application for public benefit. To this end, the Bureau conducts research and provides: (1) a basis for the Nation's physical measurement system, (2) scientific and technological services for industry and government, (3) a technical basis for equity in trade, and (4) technical services to promote public safety. The Bureau's technical work is performed by the National Measurement Laboratory, the National Engineering Laboratory, and the Institute for Computer Sciences and Technology.

THE NATIONAL MEASUREMENT LABORATORY provides the national system of physical and chemical and materials measurement; coordinates the system with measurement systems of other nations and furnishes essential services leading to accurate and uniform physical and chemical measurement throughout the Nation's scientific community, industry, and commerce; conducts materials research leading to improved methods of measurement, standards, and data on the properties of materials needed by industry, commerce, educational institutions, and Government; provides advisory and research services to other Government agencies; develops, produces, and distributes Standard Reference Materials; and provides calibration services. The Laboratory consists of the following centers:

Absolute Physical Quantities[2] — Radiation Research — Chemical Physics — Analytical Chemistry — Materials Science

THE NATIONAL ENGINEERING LABORATORY provides technology and technical services to the public and private sectors to address national needs and to solve national problems; conducts research in engineering and applied science in support of these efforts; builds and maintains competence in the necessary disciplines required to carry out this research and technical service; develops engineering data and measurement capabilities; provides engineering measurement traceability services; develops test methods and proposes engineering standards and code changes; develops and proposes new engineering practices; and develops and improves mechanisms to transfer results of its research to the ultimate user. The Laboratory consists of the following centers:

Applied Mathematics — Electronics and Electrical Engineering[2] — Manufacturing Engineering — Building Technology — Fire Research — Chemical Engineering[2]

THE INSTITUTE FOR COMPUTER SCIENCES AND TECHNOLOGY conducts research and provides scientific and technical services to aid Federal agencies in the selection, acquisition, application, and use of computer technology to improve effectiveness and economy in Government operations in accordance with Public Law 89-306 (40 U.S.C. 759), relevant Executive Orders, and other directives; carries out this mission by managing the Federal Information Processing Standards Program, developing Federal ADP standards guidelines, and managing Federal participation in ADP voluntary standardization activities; provides scientific and technological advisory services and assistance to Federal agencies; and provides the technical foundation for computer-related policies of the Federal Government. The Institute consists of the following centers:

Programming Science and Technology — Computer Systems Engineering.

[1] Headquarters and Laboratories at Gaithersburg, MD, unless otherwise noted; mailing address Washington, DC 20234.

[2] Some divisions within the center are located at Boulder, CO 80303.

Computer Science and Technology

NBS Special Publication 500-96

The Selection of Local Area Computer Networks

Robert Rosenthal, Editor

Institute for Computer Sciences and Technology
National Bureau of Standards
Washington, DC 20234

Prepared by:

Sytek, Incorporated
1153 Bordeaux Drive
Sunnyvale, CA 94086

U.S. DEPARTMENT OF COMMERCE
Malcolm Baldrige, Secretary

National Bureau of Standards
Ernest Ambler, Director

Issued November 1982

Reports on Computer Science and Technology

The National Bureau of Standards has a special responsibility within the Federal Government for computer science and technology activities. The programs of the NBS Institute for Computer Sciences and Technology are designed to provide ADP standards, guidelines, and technical advisory services to improve the effectiveness of computer utilization in the Federal sector, and to perform appropriate research and development efforts as foundation for such activities and programs. This publication series will report these NBS efforts to the Federal computer community as well as to interested specialists in the academic and private sectors. Those wishing to receive notices of publications in this series should complete and return the form at the end of this publication.

National Bureau of Standards Special Publication 500-96
Natl. Bur. Stand. (U.S.), Spec. Publ. 500-96, 133 pages (Nov. 1982)
CODEN: XNBSAV

Library of Congress Catalog Card Number: 82-600635

U.S. GOVERNMENT PRINTING OFFICE
WASHINGTON 1982

For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, DC 20402
Price $6.00
(Add 25 percent for other than U.S. mailing)

TABLE OF CONTENTS

Page

The Selection of Local Area Computer Networks

ABSTRACT

These guidelines present features available in contemporary local area computer networks including distinctions between network applications, topology, protocol architecture and transmission media. Guidance is given to identify the installation's needs. These needs are described in terms of network reliability, traffic characterizations, expected growth, and maintenance requirements.

Key words: Feature analysis; guidelines; local area networks; local network specification; requirements analysis.

1. INTRODUCTION

Local area computer network applications in the Federal Government are quite varied. An attempt to meet the needs of all applications would produce a document too general to be of use to its readers. Therefore, these guidelines are oriented toward computer based office system applications where information transmitted may be data, voice, and/or video, and the equipment employed may be interactive terminals; printers; micro-, mini-, and/or mainframe computers; word processing stations, and other similar devices. Users perform their jobs in an interactive mode making use of on-line facilities such as compilers, data base management systems, utilities and other available application programs. In addition, advanced communication services such as electronic mail, teleconferencing and facsimile may be required.

This report does not attempt to precisely define local area computer networks. In general, these networks are deployed in small geographic areas such as an office complex, building or campus. Typically, they are owned, operated and managed locally rather than by a common carrier. Bus, ring, star and mesh topologies that support the interconnection of (up-to) hundreds of devices communicating by packet transmissions at rates in the 1 to 20 megabit per second range are not unusual.

1.1 Technical Approach

The selection process consists of the following steps:

1. Perform a technology and features survey of commercially available local networks. A "Feature Analysis of Local Area Computer Networks" [GARL81] supplements these guidelines by analyzing local network functions and capabilities in terms of services provided, topology, protocol architecture and transmission medium.

2. Perform an analysis of the installation's requirements. A "Requirements Analysis of Local Computer Networks" [ENNG81a] supplements these guidelines by enabling ADP managers to determine requirements in five areas: network service, network traffic characterization, network reliability, network growth and network maintenance.

3. Prepare a network solicitation using the requirements identified in step #2. A "Specification of Functional Requirements for Local Area Computer Networks" [ENNG81b] supplements these guidelines providing a step-by-step procedure to follow with suggested metrics in the areas of services, traffic, reliability, growth and maintenance.

4. Evaluate responsive proposals.

This report is derived from the feature analysis, requirements analysis and specification of functional requirements referenced above.

1.2 Report Outline

This report is presented in seven sections describing the steps that a data processing manager might follow in the pre-solicitation analysis, the preparation of a solicitation document and the evaluation of proposals.

Section 1 is this introduction.

Section 2 presents features available in contemporary local area computer networks, including distinctions between network applications, network topology, protocol architectures, and transmission media.

Section 3 provides guidance for the requirements analysis. It identifies the range of services necessary to meet the needs of the target installation, the traffic characteristics of network nodes, reliability, expected growth, and maintenance.

Section 4 describes the specification of requirements for the network solicitation document.

Section 5 describes a quantitative method that may be used to evaluate responses and to select the local network best suited to the installation requirements.

Section 6 presents the summary and conclusion to this report.

Section 7 lists the references used in the report.

2. CONTEMPORARY LOCAL AREA COMPUTER NETWORKS

A variety of alternative local area network designs exist. To compare these alternatives, local network features are identified. One local network is distinguished from another on the basis of:

1. intended application and services offered,

2. network topology,

3. protocol architecture, and

4. underlying transmission medium.

The approach is to analyze local network features using these four distinguishing characteristics. Limitations imposed by these features and design tradeoffs made are included under each topic.

2.1 Application Distinctions

Networks combine transmission, storage, and processing operations to meet user application requirements. Typically, a local network provides these operations for specific application types. In this section four application types are described: 1) the type of information transmitted such as data and/or voice; 2) the type of interacting devices such as terminals and computers; 3) the type of service provided such as computer based message systems or electronic mail; and, 4) specialization of the network environment such as process control or laboratory automation.

2.1.1 Data, Voice, and Data/Voice Networks. The type of information transmitted has varying characteristics. Consider voice and data. Typically, voice differs from data in error and buffer handling techniques and in bandwidth requirements. While other information types such as imagery or video may be transmitted through local networks, data and voice provide an appropriate comparison point.

Although always digital in nature, the characteristics of data transmission widely vary. Two types of communication contribute to typical traffic patterns:

1. Interactive users at CRT terminals to/from host computers, and

2. Computers to/from other computers or intelligent devices.

In the first case, terminals operating at data rates in the 9.6 thousand to 19.2 thousand bits-per-second range offer bursty traffic to the network. In the second case, computers offer higher data rates. Often computer-to-computer data rates appear in bursts due to protocol overhead and the scheduling of host processing cycles. In general, transmission of data between source and destination must be timely and error-free, usually requiring retransmission when errors are detected.

The telephone call and connection is the common unit of traffic in voice communications. The mean and standard deviation of call hold times characterize the traffic. Typically, the average calling rate on a telephone line is a few per hour and the mean call hold time is two to three minutes.

The concept of a call is appropriate to systems where a connection is made between users, held for a substantial period of time, and then broken. Implementations of voice communications have historically resulted in "circuits", where once a connection is made, a portion of the transmission medium becomes dedicated to that circuit until a disconnect takes place. Information is transmitted without interruption in real-time so that individuals on both ends continue to participate in a natural fashion; up to a point, transmission errors can be tolerated.

Voice-oriented networks historically have been designed using a variety of circuit-switching techniques. These networks are wasteful of circuit resources when modified to handle data.

Data-oriented networks are designed assuming computer and peripheral device interactions are transaction oriented. Individual peripheral devices typically generate a small percentage of the overall transactions. Information in these transactions is delivered in a timely manner and error-free. Data networks that carry voice must be augmented to handle these conflicting characteristics.

Integration of heterogeneous traffic into a common network is desirable for economic reasons and for simplicity of operation. Integration provides for the dynamic sharing of transmission and switching facilities and may encourage new applications such as teleconferencing. Networks that support both voice and data can be designed; a variety of switching technologies have been investigated for such integration taking into account circuit- and packet-switching concepts [GITI78].

2.1.2 Terminal and Computer Communications. Terminal-to-computer and computer-to-computer applications differ in the speed, frequency of use, and service provided for device to device communication.

Terminal to computer communications are typically transaction-oriented or "bursty", and include data entry, word processing, program development, data base access, and remote job entry. Applications require conversational network services at speeds slower than computer capabilities. Interactions are infrequent, with maximum effective terminal speeds of up to of 19.2k bits-per-second.

The primary objective of these applications is to provide geographically distributed user terminals and other types of devices with remotely located data bases and/or computing power. As a consequence, there exist two types of network node: those that represent terminal entry/exit points to the network and those that represent processing and data base points to the network.

Applications in which computer-to-computer communications take place are high speed; they consisting of:

1. File transfer and,

2. Distributed processing, including inter-process communications.

Computer-to-computer interactions take place at the maximum speeds allowed by the computer and network architectures, with speeds of up to 50 M bits/second. The applications are oriented toward the same types of communication found between processes in a single computer system. These include the exchange of files and inter-process messages.

Network design and implementations are often oriented toward the highest volume of traffic required by the target installation. However, most networks require support for both types of traffic; and this implies that compromises and design trade-offs are made.

Terminal networks are generally too slow for the response required by computer-to-computer traffic. Equipment necessary to provide the high computer-to-computer speeds is more expensive then that for slower traffic types. The cost of the interface node to the network is significantly less than the device being connected.

2.1.3 Advanced User Services. Office automation refers to the aggregation and integration of several applications of computers and networks in office work that aid the completion of normal office functions. Office automation includes everything presently called "word processing" (dictation, document preparation and editing) plus computer-based filing (information storage and retrieval), communication (electronic mail, teleconferencing), and modeling (simulation), interacting with management information systems and other functions.

Heavy use of local and geographically distributed networks will be made in the distributed automated office environment. Much office work is organized in a hierarchical manner, with component desk functions such as transcription, editing, filing, retrieval, scheduling, and telephone answering at low echelons with agency or divisional functions such as planning and operations at high echelons. Low echelon functions are typically carried out locally within a single office or suite of offices, supported by minicomputers or microcomputers. Geographically distributed networks may be required as higher echelon functions are integrated.

Simple message communication and file transfer, access to time-shared processing and access to information are fundamental network support functions. More advanced user services that are highly differentiated and specialized make use of these functions. The functions can be either external, where they are visible and known to network users, or internal, where they are required for network operation and maintenance.

Some external services include:

Data Base Access:

Management, professional, and non-technical support personnel require cost-effective and efficient access to an extensive amount of information. The local network provides appropriate services and user interfaces. Responsibility for processing user data base requests is divided between user/network interface devices and data-base specific software in specialized distributed or centralized data base nodes.

Centralized File Access:

An important application is file access and storage where little or no data interpretation is performed by a centralized file storage node. The communication requirements of centralized file access are distinct from connectionless applications, in that connection-oriented services are necessary for reliable delivery of sequenced data. File transfer capabilities are required [CLOS80a] [CLOS80b] with translations among the file system conventions of different machines. File archiving, with network access to a shared long-term storage device and an appropriate user interface, is often a primary requirement.

Message Processing and Electronic Mail:

Message processing facilities are examples of local network applications with high user visibility. As the size of messages vary depending on user requirements, network services typically support low volume connectionless and connection-oriented (file transfer) traffic with store and forward facilities. Editing, filing, and formatting support to users is desirable as identified in the proposed ICST Message Format Standard [BOLB81].

Bulk Raster Traffic:

Services facilitating the exchange of facsimile and digital voice messaging applications place requirements on the network for the transmission of bulk raster traffic. Connection-oriented transport services with wide bandwidth capabilities are appropriate. Translations among the various codes and transmission formats found in the facsimile industry may be necessary.

Graphics:

Coded raster information substantially reduces the required bandwidth for transmission of graphical data. The coded nature of graphical data requires reliable, in-sequence delivery since information loss or rearrangement affects the received image.

Real-Time Raster Traffic:

Real-time voice communications and slow or medium scan television demand network performance that guarantees specific throughput rates, response times, and high bandwidth. The combination of such communications into an integrated service requires facilities which coordinate multimedia and multipoint transmissions.

Teleconferencing:

Communication requirements imposed by teleconferencing applications demand real-time transport and synchronization of multiple data streams (audio, visual, graphical, and textual) between participants. Both distributed and centralized implementations are feasible.

Internal functions fit into the following categories:

Directory Services:

When a large user population exists, the desirability of augmented facilities for addressing users is suggested. The ability to provide addresses in the form of symbolic names rather than numeric strings is both a cost and work-saving improvement. A technique adopted in many local networks is to provide a directory service that translates symbolic network user/subscriber names into internal network numeric addresses.

Billing and Accounting:

Networks are often utilized by numerous groups, for which billing and accounting may be necessary. Network services may not be free; and billing may be based on usage. Internal mechanisms may be required for the collection of data describing each subscriber's usage, source and destination of the service, length of service, and amount of data transported during the service.

Status and Performance Monitoring:

The operation of any network with geographically distributed equipment requires collection of equipment status information and evaluation of overall network performance, in terms of error rate, utilization, throughput, delay fairness and stability. Provisions for the collection and evaluation of this information for all network services should be included in the protocol architecture, where status and performance monitoring functions are required to guarantee quality of service throughout all protocol layers.

Privacy and Security:

A data network's security mechanism protects users from misdelivery of data and provides for user authorization and protection of data. Misdelivery could result from network error or deliberate penetration upon the network.

Access Authorization and Protection:

Internal means for user identification and for restriction of access to specific network services to differing user groups is necessary in networks where billing or accounting takes place. Legitimate network subscribers should not be able to improperly access data; this requires the authentication of users who attempt to access network or user resources.

2.1.4 Specialized Networks. Two specialized local network applications of interest to federal agencies are process control applications and laboratory automation applications.

Machine and process control applications for computer systems have been in existence for a number of years. Early process control minicomputers served as monitors repeatedly measuring process variables, comparing measured values to preestablished limits, and sounding alarms if out-of-limit conditions were discovered. Systems control has become more sophisticated over time, where processes can now control the machining of parts from engineering drawings, for example, as well as perform monitoring functions. Management and cost accounting functions are also performed utilizing interactive terminals at various locations throughout a manufacturing facility. These same terminals are used to follow particular jobs throughout the manufacturing process; data may be employed to analyze utilization, reliability, and scheduling.

Networks that support process control environments include capabilities for data transmission in real-time with minimal and predictable delay. Also, the networks are usually capable of handling lower-speed interactive (terminal) data transmissions concurrently with the real-time data.

The laboratory and instrumentation environment is experimental, with equipment configurations and applications constantly changing as the experiments and tests evolve. Very specialized and expensive equipment is often included, using interfaces such as IEEE-488 [IEEE78] and CAMAC [IEEE76a] [IEEE76b]. Minicomputers are an invaluable part of such systems, providing program-controlled instrument set-up and data acquisition. Central computer systems are often used for data reduction and analysis, data and program storage, and special services such as data plotting.

Networks utilized for the automation of this environment are capable of supporting a variety of parallel and serial interfaces and of handling the dynamics of growth and change. Growth is defined in terms of additional nodes, increased traffic, and increased geographic scope; change implies differing types of traffic and interfaces, and mobility of equipment location.

2.2 Network Topology

Network topology determines the manner in which the switching nodes, user devices, and transmission links are inter-connected. There are four basic topologies, with most actual networks designed using a mixture of topology types. The following sections describe star, ring, bus, and mesh topologies, with their relative advantages and disadvantages compared in terms of reliability, expandability, and performance. Performance is discussed in terms of delay-throughput characteristics of the network topologies, where delay is the mean transfer time of transmitted information blocks, or the time interval from the generation of a block at the source until its reception at the destination.

<u>2.2.1 Star Topologies.</u> In a star topology network, each network node accepting and delivering user traffic is connected to a single central node through which all traffic must pass. The central node acts as the system control, with separate lines to all other nodes. Usual implementations result in system control performed by a central switch, with all participating network nodes connected to that switch by dedicated lines. When messages are sent between network nodes, the transmitting node makes the request to the central switch, which in turn establishes a path to the receiving node.

Star topology networks are of two types: circuit switched or message switched.

Circuit switching is a method of communications in which a connection between calling and called nodes is established on demand for exclusive use of the circuit until the connection is released. Modern PABX (Private Automated Branch Exchange) systems that carry both voice and data are circuit switched networks.

Message switching is a method of handling messages in networks where the entire message is transmitted to an intermediate point (i.e. switching computer), stored for a period of time (usually short), and then transmitted again toward its destination. The destination of each message is indicated by an address field in the message. Store-and-forward processing is typical of contemporary distributed processing systems made up of a number of stars, where the central computer of one star has the capability to communicate with central computers of other stars often located some distance away.

Single-node failures present minimal reliability problems, since the remaining portion of the network is still able to function. Failure of the central switch, unless implemented with full redundancy, brings the entire network to a halt.

A single star configuration may be expanded up to the limitations of the central switch. Configurations grow in terms of switching capacity (number of switches that can be made in a given amount of time), numbers of concurrent circuits that may be maintained, and total number of nodes that may be serviced. Switching capacity depends upon the message rate, throughput and the processing time required for each message.

When further expansion is required, the star topology becomes a tree-like structure, where points of the star become centers of additional star structures performing switching functions. Clearly the expansion can grow to multiple levels. Without redundancy, failure of the central switches at each level present reliability problems within the star and at the levels below.

Throughput in a star topology network is the time required for the central switch to process and route a message. Message processing time depends upon the number of network nodes and message length. The throughput rate is the reciprocal of the sum of processing time and input and output buffering times. Using software implemented switches, current PABX systems typically handle 2000-4000 concurrent circuits, with devices communicating at effective speeds of up to 56 K bits/second.

<u>2.2.2 Ring Topologies.</u> A ring network consists of nodes with connections only to the node on each side, such that a complete circle is obtained; the design is an attempt to avoid the potential reliability problems with the central node of a star. Nodes are able to transmit or receive data in either direction on the ring with full duplex links, but data must pass through all nodes between the sender and the receiver with the possibility of a shorter path in one direction or the other. Unidirectional ring topologies are more common. When messages are sent from one node to any other they are entered onto the ring and travel around it until received at the addressed node or returned to the transmitted node; the message may be removed by the destination node or the originating node, depending upon implementation and protocol.

When messages are removed from the ring by the originating node, that node has the opportunity to perform an error check by comparing the original transmitted message with the returned message.

Unidirectional ring topologies have the following attributes:

1. Message routing is simplified, since only one message path is possible, considering the most common implementation.

2. A transmitting node need only know an identifier for the receiving node, and not its location.

3. Broadcast message transmission is easy to achieve, since every node is capable of receiving the message as long as previous nodes have not removed it from the ring.

4. There is the possibility for low capital investment, with cost proportional to the number of users or interfaces.

5. The possibility for high throughput exists since more than one message can be in transit at once.

Ring topologies are categorized based upon the type of message transmission mechanism employed. Examples of transmission mechanisms are:

1. The round-robin passing of a control token from one node to the next, where only the node with the token can transmit an arbitrary length message (Newhall-type loop [REAC75]);

2. The slotted ring in which fixed-sized slots are cycled around the ring; with a bit indicating whether the slot is full or empty, the node can fill a slot as required (demand multiplexing) or can be assigned a specific slot or set of slots (synchronous time division multiplexing). Slotted rings are also known as Pierce-type loops [OHY77];

In any ring control strategy there is some entity, be it a control token or an explicit signal on a wire, that is passed from node to node to indicate which node currently has the right to transmit. The strategy takes into consideration the possibility that a transient error may destroy the entity, and therefore, the entity must be prepared to restart itself after such an error. This is accomplished by regenerating the permission to send and bestowing it uniquely upon one of the nodes. While it is difficult to determine with certainty that the control entity has been lost and to decide which node should take it upon itself to recreate it, solutions do exist; usually, these solutions include some form of contention.

Transmission and processing errors can affect proper network operation as in the following example. If a message is somehow distorted with the address modified, the message could be delivered to the wrong destination. An invalid address, if not properly handled by the system may result in a message that continues to circulate around the ring. Many systems use varying monitoring schemes to check and remove messages that travel without being received.

Since nodes must be capable of selectively removing a message or passing it on to the next node, they each must contain an active repeater. Single or multiple node failure could either cause a loss of access to the ring or a breakdown in total operation. Often, ring implementations include two electrically independent devices: 1) a simple primary section that is line-powered and 2) an interface section that is relay operated and locally powered. If a failure occurs the relay circuit decouples the bad component maintaining proper ring operation.

A number of improvements to the ring topology have been proposed and implemented [HOLK76] [HAFE76] [ZAFP73].

Expansion is usually simple. It requires the insertion (electrical) of the new node's interface into the ring. It also involves identifying the new address to the neighboring nodes. Expansion may have a distinct effect on network performance since all messages pass through the new node.

Performance of a ring network is dependent upon the message transmission mechanism. A simplistic mechanism may require each node to take messages into buffers as they are passed from node to node in order to interrogate an address field. If heavy traffic conditions exist, messages may have to wait in long queues at each node for input and/or output. Improved methods exist such as Delay-Insertion [MAFE74], where delay can be reduced to as low as one bit time per node. Average transmission time on a loop is independent of traffic load for token and slotted rings; in delay-insertion the mean transmission time increases significantly with higher traffic loads, but is superior where queuing delays for messages entering the ring are short [REAC75] [OHY77] [MAFE74].

Rings provide the best performance for networks with a small number of nodes operating at high speeds over short distances. The token ring [BUXW80] in general performs quite well, as long as implementations result in low delay at each station. Backbone rates of a hundred megabits per second and more are achievable.

2.2.3 Bus Topologies. Network nodes connect to the same transmission medium in bus topologies. Each node has an address used to uniquely identify the recipient of a transmitted message. The bus is typically time or frequency multiplexed, allowing nodes to transmit information in short duration, high-speed bursts.

There are a variety of mechanisms that arbitrate use of the bus. These mechanisms are either centralized or distributed. Centralized control schemes concentrate hardware in one location. All messages are first transmitted to a switch that retransmits the messages over the same bus to the proper destination. The control can be within one node or it can be performed by a dedicated bus controller that operates in one of two ways:

Polling by controller mode, where the dedicated controller polls each of the nodes connected to the bus for information to transmit.

Device interrupt to controller mode, where each node may request one or more slots for transmission or may simply request that a message be transmitted to its destination.

In a decentralized control environment, bus control logic is distributed throughout all the nodes connected to the bus. This technique is based upon the notion of a global bus, and is the most commonly used technique. Variations to the global bus include:

Frequency division multiplexed (FDM) bus, where each receiver and transmitter associated with a node is assigned a distinct frequency.

A variety of time division multiplexing (TDM) techniques are discussed. They are: fixed slotted, dynamic slotted, ALOHA, slotted ALOHA, CSMA, CSMA/CD, Reservation TDMA, and token bus [CLAD78] [SHOJ79] [KLEL76] [LAMS80]. Each of these techniques is useful on different transmission medium. Though it is the most inefficient of the access methods, ALOHA is particularly well suited to radio transmission, where listen while talk strategies will not work because the transmitter overloads the receiver. Slotted techniques more efficiently use the available bandwidth capacity and are useful for full-duplex cable bus media where synchronization information may be received concurrently with transmission/reception [KAHR78]. CSMA and CSMA/CD are more efficient and better suited for networks with many devices infrequently accessing the network. Good utilization of the cable bus media for networks supporting devices with frequent high-volume accesses can be obtained from Reservation TDMA access methods [KLEL76] [LAMS80].

A bus is a passive medium by nature, with each node listening to signals in transmission. A node can fail without disrupting the bus if it fails in a manner that presents a high impedance to the bus. A watch-dog timer on the transmitter of each node should detect and disconnect a node that fails in the transmit mode. The failure of active bus components normally disables nodes beyond the failed component unless erroneous signals are introduced that interfere with information signals. The location of the active component will determine how much of the network is affected. Redundant components for each active component with automatic switchover capability can be used for high reliability networks. Subjecting the transmission media to a catastrophic disruption, such as lightning or an errant cross connection to power lines, may destroy electronic components connected to the medium. The susceptibility of the cable bus network to this type of failure is dependent on the isolation of the cable connections and fusing of the active components.

The reliability of a bus network depends also upon the topology and control strategy as an intertwined entity. When using contention control, a very appealing aspect is that almost any failure has exactly the same effect as a collision, and is thus dealt with automatically. If a message is garbled it must be retransmitted, but no long-term failure of the network results. Contention control does require that the recipient detect a garbled message and be able to request a retransmission if the original transmitter has for some reason failed to discover that the message was garbled. Higher level protocols must provide mechanisms that ensure reliable recovery. Note that attention to reliability at higher levels is required regardless of the control strategy of the network, and is not any particular disadvantage of contention.

Bus structure expandability is primarily based upon the transmission system utilized. Expansion may require the inclusion of amplifiers and/or repeaters at periodic locations in the medium to ensure signal quality, up to the limits of the transmission hardware and the networking technologies employed for software and hardware. Often the busses are easily tapped, allowing the inclusion of new devices and/or bus segments at any bus location.

Similar to other topologies, bus network performance is determined by bus bandwidth, number of nodes connected to the bus, bus access protocol, timing, average and peak bus user traffic rates, and other tolerances introduced by the design approach taken. Using contention control schemes, for example, utilization of the bus may be as high as 95%, but performance will drop as the frame sizes decrease and as propagation delay between the farthest network nodes increases. Ideal conditions require a low ratio (< 5%) of propagation delay to mean transmission time [BUXW80]. Single channel backbone rates with high utilization are achievable to the 50 megabits per second range, for small numbers of devices communicating over very high quality cable up to 1500 meters in length.

2.2.4 Mesh Topologies. Mesh topologies involve redundancy of transmission paths between communicating nodes, with a duplicate path between each pair available (see Figure 2.2.4-1). Such topologies are basically point-to-point partially or fully interconnected systems, where the interconnect complexity can grow rapidly as the number of nodes connected to the system increases. For this reason partially interconnected systems are more common, using transmission switching schemes to route data through the network. Network node location is usually dictated by user need, cost considerations, performance, and anticipated network traffic. Utilization of mesh topology is typical for large, long-haul, packet-switching networks.

The generality of this arrangement introduces the unavoidable cost of making a routing decision at each node of the network, but is capable of adaptation to changing traffic patterns. Each link must decide after receiving a message to which link it must be forwarded if that link is not the intended destination; this implies substantial computations at each node [CLAD78]. Few local networks have implemented this generalized structure, due to the intelligence complexity and resultant cost of each node; however some implementations of local networks do make use of mesh related path control capabilities for load leveling.

Mesh topologies have similar reliability problems to those of ring networks, but not to as high a degree. The possibility for error exists as messages are passed between nodes on the computed route to the destination, with intelligent processing performed at each step and possible inadvertent message modification. Single node failures do not present extreme reliability problems in that alternate paths should always be available unless other nodes have also failed; performance may be affected, however.

Expansion of a mesh topology network to include additional nodes is straightforward as long as the two or more other nodes with which the new node is to communicate can handle the increased load. Considering the loading detection and routing decision strategy utilized, software table structures required for routing decisions in all nodes may need update for inclusion of the new node. Performance may be degraded if the additional nodes also cause additional hops before reaching the final destination.

Performance of a mesh network utilizing adaptive routing techniques is difficult to predict, where the prediction is highly dependent upon the complexity of the routing algorithm and the accuracy in time and detail to which loading statistics are kept. Methods for keeping and using those statistics are difficult problems themselves [DAV79]. The problem of performance prediction can be tackled by simulation. Details of the routing algorithm have distinct effects on performance [KERI76].

Bus networks using gateways to interconnect local networks within a close locality of reference via fixed addressing can be logically considered as mesh networks. Performance measures between nodes at different networks communicating via gateways are difficult to predict due to the same considerations as above. Performance within a local network employing gateways is quite similar to that of bus networks, where additional delay is imposed by the gateway unit to provide frame buffering between local networks. In order to optimize performance, networks so implemented should be careful to place communicating devices on the same network to avoid gateway interfaces where possible.

2.3 Protocol Architecture

Communication protocols are required to effectively manage the use of network hardware and to provide basic services. For example, protocols are necessary for error control, maintenance of proper data sequencing, data transfer, routing, and congestion control.

The International Standards Organization (ISO) has defined a Reference Model for Open Systems Interconnection that specifies an architecture for open systems communication. This model, used as a reference in comparing different protocol architectures [ISO82], allocates functions to protocol layers.

2.3.1 The ISO Reference Model. The International Organization for Standardization (ISO) has developed a Reference Model for Open Systems Interconnection (OSI) [ISO82] that specifies an architecture for open systems communications.

The layered ISO Reference Model provides a well defined environment for protocol development. It defines how seven layers provide services that support communications among application processes in an open system. Higher layers use the services provided by the lower layers; the layers themselves communicate on a peer basis: layer 1 to its peer layer 1 and so forth.

2.3.2 Services Offered. The services offered by each layer are described below. These are services as indicated in the Reference Model. All of them do presently occur the ISO protocol work.

1. Physical Layer

Arbitration of access to the transmission media is handled by the physical layer. Examples of several local area network access schemes are given:

ALOHA

The pure ALOHA scheme was originally developed for satellite packet radio using a broadcast radio channel as the shared communications medium. The scheme requires that a node start a random timer when a message is transmitted, and if the message is not acknowledged before the timer runs out, it is retransmitted. A message can be transmitted at any time.

The technique provides a throughput of about 18% of the circuit capacity. At high loads (greater than or equal to 18% of the capacity) the method becomes unstable [KLEL76].

SLOTTED ALOHA

SLOTTED ALOHA provides increased performance over ALOHA. Transmissions can only begin at specified time slots. The maximum capacity of the network is doubled to about 36% throughput [KLEL76].

TDMA

Reservation Time Division Multiplex Access (TDMA) uses a conventional TDMA structure with a dynamic scheme for the allocation of transmission slots. This mechanism works well with high duty cycle Data Terminal Equipment (DTE), substantially less well with low duty cycle, bursty devices. The internal protocol required for reservation of bandwidth is complex in addition to the hardware required for TDMA frame and slot construction and derivation. Data fragmentation due to a mismatch between DTE data frame and TDMA slot size also introduces complexity.

CSMA/CD

Carrier Sense Multiple Access (CSMA) and CSMA with Collision Detection (CD), requires that nodes listen before transmitting. If another node is transmitting, the listening node will either back off for a specified time interval before listening again or continuously monitor until the network is clear to send. Collision Detection includes a mechanism that if a collision occurs during transmission, both sending nodes back off random time intervals before trying again. For small networks, the maximum capacity CSMA can provide is roughly 85% [KLEL76]. This value decreases with increasing propagation delay relative to frame length. At high loads the algorithm exhibits some instability. CSMA/CD [TOBF79] increases maximum utilization to 98% of the inherent circuit capacity and exhibits stability even in the presence of extreme overload [SHOJ79]. Performance decreases with decreasing frame size and increasing propagation delay between the farthest transmitters. An attraction of CSMA/CD is the simplicity of its implementation and the inherent system reliability.

TOKEN RING

In token ring systems a control token is passed sequentially around the ring, where the node with the token is allowed to transmit. Assuming that the delay time per station is small, the capacity of the token ring approaches capacities equivalent to CSMA/CD (~98%) [BUXW80]. For higher speed networks (10 Mbs or greater), in which the ratio of the propagation delay to packet-transmission time gets too high, the capacity of the CSMA/CD drops significantly (66%), while the token ring capacity remains at 98% [BUXW80]. There are delays caused within stations by token handling, address recognition, or altering of control information. These delays can become significant if there are a large number of stations or packet sizes are small. Transmission delays two to three times greater than those for CSMA/CD can result if these station delays are large (8 to 16 bit times) [BUXW80].

2. Data Link Layer

Management of the communications link at the physical access layer includes allocation of and contention for the transmission medium. The second layer of the ISO Reference Model describes the need for node-to-node protocols more commonly referred to as data link control procedures. These protocols govern information flow across a physical circuit.

Quality of service parameters for the data link layer may be optionally selected, established, maintained for the duration of the data link connection. Quality of service parameters may include:

1. Throughput - information transfer capacity of the link,

2. Transit delay - maximum delay allowed,

3. Service availability - allow or prevent the blocking of services to a link,

4. Residual error rate from loss, duplication, disordering, etc.,

5. Detected error rate - error detection algorithm selection, and

6. Delay variance - the maximum allowable variation in delay from an established mean value.

The link layer protocols are concerned with the actual transmission of bits or bytes over the communication link, error detection and correction, information coding, information transparency, line utilization, synchronization, communication facility transparency, and bootstrapping.

The primary function of data link control is to manage the transfer of information across an established data link. Data transfers are controlled using three different elements: formatting, control information, and handshaking procedures. Formatting means reserving positions, or fields, in the transmission block for specific information. The data link layer protocol delimits fields with control characters. In byte or character-oriented protocols, control characters are distinguished from and cannot be used as information characters (except in conjunction with other control characters). To control the flow of information, the control field usually contains addressing, block sequencing, control flags, and acknowledgement information. Handshaking covers the establishment and disestablishment of the data link. Link establishment includes all functions that must be completed before both ends of the communication link are satisfied that communications are possible and can proceed. Link termination is the orderly disestablishment of the data link, with procedures that provide data link closure information to both ends of the communication link.

A second type of data link service, known as connectionless service or datagrams, can be used in place of the connection-oriented service, and does not require the handshaking procedures described above. A connectionless service is not currently described at the Data Link Layer of the ISO Reference Model, though it is anticipated that it will be added in future extensions of the Reference Model. This is described as Class I service in the IEEE Computer Society's Local Network Standards Committee Project 802 Data Link Control Protocol [IEEE80]. In the connectionless service a single unit of information, a frame, can be passed from one node to another. In most existing versions of the data link layer this service is provided without guarantee or acknowledgement of delivery, though error control and reliable delivery can be offered as quality of service additions. Connectionless service is the basic service type of most local networks because it takes advantage of the basic characteristics of local networks (i.e. low delay, high throughput, and low errors). The use of connectionless service is important because it prevents inefficiency in those applications which do not require reliability guarantees. Packetized voice, data enquiry, and status transmissions are applications which do not require 100% reliability at the data link layer [CALR81b].

An important function of a data link protocol is to assure correct reception of data because communication facilities are error prone. To provide this, link control procedures include the generation, transmission, and testing of check bits, generated by one of the following checking algorithms:

1. Odd or even parity checking per character (VRC),

2. Block checking using exclusive "OR" logic (LRC), or

3. Cyclic Redundancy Checking (CRC) - block checking using a polynomial division of the data stream by a CRC polynomial.

The CRC method is a more sophisticated approach. Quality of service parameters can specify the quality of error detection provided; this service can be implemented using more complex detection algorithms. Error detection service can be provided for connection-oriented or connectionless service by using the frame CRC. The error detection on connectionless service would be useful for status and control applications that required low delay but reliable delivery.

Information transparency must be provided by the connection or datagram data link protocol to differentiate control information from data. The move from byte-oriented to bit-oriented data link protocols was in part prompted by the restrictions placed on byte-oriented protocols in which certain characters are reserved as control characters and so can not appear in the data field without special delimiters.

Bit-oriented data link protocols permit the use of all bit patterns in an information field and still control the transmission of the block. This transparency facilitates the transmission of binary data, floating point numbers, packed decimal data, unique specialized codes, or machine language computer programs. Transparency is normally accomplished in bit-oriented protocols by "bit stuffing/unstuffing". Under this strategy, all bits in the information field are scanned to be sure that the bit patterns similar to control bit patterns are eliminated.

The data link layer can assure the sequence of delivery of frames. This is usually done by uniquely numbering the frames in some agreed upon sequential order. Normally a modulo based number scheme is used to limit the number of bits required to represent the sequence number(e.g. 8, 128, etc.). The sequence number of each frame received is checked to verify that the frames arrive in sequence. Out-of-sequence frames are marked as in error with correction mechanisms employed.

Sequencing at the data link layer normally applies to connection-oriented protocols only. Datagrams (connectionless service) do not provide sequencing at this layer, though higher-layer protocols may have to provide sequencing to support fragmentation required at those layers. This is because datagrams are considered atomic units or transactions that have no relationship to preceding or subsequent datagrams.

Flow control is the set of mechanisms where a flow of data can be maintained within limits compatible with the amount of resources available [POUL80]. The resources required may include buffer space, transmission bandwidth, name space, table entry space, logical channels, or process time.

Each node can dynamically control the rate at which it receives frames from a data link connection. This rate of flow can be controlled by:

- a simple start-stop mechanism by which the receiving node indicates when it is ready for the next frame,

- a window mechanism in which the receiving node indicates the number of frames it will receive (can use the frame sequence number), or

- As with sequencing, datagrams at the data link layer do not normally provide for flow control [IEEE80].

- Use of connections, datagrams, or both at the data link layer depends on factors such as the quality of service parameters required by higher layer protocols, and the media and access requirements and constraints.

- Specific networks may be required to support voice applications. These applications may dictate the use of a connection service to meet the special flow control, speed, and delay requirements. Other applications such as file transfers, real-time, interactive, and electronic mail delivery services may use datagrams or virtual connections built on the datagram service.

- Higher layer protocols may require quality of service support at the data link layer that dictate the use of a connection or datagram service. Requirements to vary the throughput, transit delay, service availability, residual or detected error rate will help identify the correct data link service.

- Physical layer constraints that may dictate a particular data link service are the reliability, error rate, and most efficient access method to the media. Unreliable media may require the use of a connection service to provide an acceptable level of reliability. Error prone media may require more stringent error detection at the data link layer. While a particular access method may work more efficiently with either a datagram service or a connection service.

3. Network Layer

The third layer of service provided is network routing and switching. The network routing service is used to get a unit of information (normally called a packet) from one network node to another regardless of whether these nodes exist on the same or different networks. A local area network that was not required to communicate with any other network, and whose topology did not require routing information, could effectively eliminate the network routing function. In the current era of ever expanding communication capabilities and requirements, isolated local networks are becoming a rarity.

The network layer is subdivided into two sublayers, one supporting network protocols for the local network referred to as the communications service sublayer and one supporting internetwork protocols (referred to by ICST as the internet protocol) [BOLB80] [CALR81a].

The network layer, like the data link layer, provides either connection-oriented or connectionless (datagram) type services. But unlike the data link layer the network layer normally does not provide for data reliability, so protocols at upper and lower layers are assumed to provide reliability. Network encapsulation could include additional error control procedures for individual high loss rate or high reliability application networks if required. The choice of service type will affect the services provided to higher-layer protocols and the promulgation of quality of services to lower protocols.

Connectionless service transports independent packets and so allows independent routing and delivery of each packet. Connectionless service simplifies internet gateways by not requiring the saving of connection state information. (A "gateway" is the collection of hardware and software required to effect the interconnection of two or more data networks, enabling the passage of user data from one to another [CERV78].)

Connection-oriented service allows the establishment of logical connections between source and destination hosts on the local network or on interconnected networks. Gateways must keep the routing and state information for established logical connections.

The greatest advantage of the connectionless service is simplicity, particularly in gateways [BOLB80]. Because each packet is treated independently, hosts and gateways do not need to maintain state information about logical connections, which in turn reduces code complexity and storage requirements. If the internet protocol is a connectionless service and the lower layer protocols are connection-oriented, then full advantage may not be taken of high reliability, sequenced delivery services provided by the lower protocols. The support of broadcast and multicast is easier with the connectionless service and is not currently available with connection-oriented services. Another advantage for connectionless service at the network layer is the ease of dynamic routing and network reconfiguration, since establishment procedures need not be executed before data transfer.

The advantages of connection-oriented services include the ability to provide sequenced, efficient flow controlled delivery. The connection service provides a straightforward interface for existing communication protocols that assume an underlying reliable delivery protocol, such as level 3 of X.25 [CCIT77b].

Quality of service parameters that can be applied at the network layer are:

1. Priority - allows the user to specify the priority of each packet for internal routing in the local network and internet processing at the gateway,

2. Cost - user can indicate that the packet should take the least expensive route possible, where the cost is computed at each gateway,

3. Delay - user can specify the maximum allowable delay for a packet,

4. Security - the user may specify security/user group information on a per packet basis,

5. Lifetime - the user may specify the lifetime of each packet within the local or interconnected networks, and

6. Grade of service - the user may specify acceptable error and loss levels or speed versus reliability preference.

Network routing is the major function performed by the network protocol. Other functions provided by the network protocol include network naming, addressing, packet formatting, and internetwork fragmentation and reassembly.

Local networks exist in which the topology allows or requires that routing decisions be made (i.e. rings may need direction of travel specified, meshes require node routing information, and broadband busses may support channel routing). These routing decisions can be tied to quality of service parameters such as cost, throughput, delay, etc., where routing decisions can be used to specify a level of service or optimization of service.

When there is a requirement to share data between networks, then an internet routing function must be used to move data from one network to another. These networks may be local area networks, long-haul packet-switched networks, or packet satellite networks providing connection and/or datagram support. This interconnected set of networks is known as a "catenet" [BOLB80]. The major consideration in interconnecting networks into a catenet is minimization of changes required to existing networks.

Unless the networks to be connected are identical such that they can be connected at the packet layer, an interconnection strategy must be defined to allow communications. This interconnection strategy can either translate the protocols between the networks to provide application to application communication, or can require that a common set of higher layer protocols be used to pass through the catenet. The second strategy, known as "interconnecting at the host layer", then provides a uniform structure for communicating over a nonuniform catenet. The use of a uniform set of internet protocols does not mean that a different set of higher-layer protocols, optimized for the local application, can not be used for local communications. Though the costs associated with maintaining a nonstandard set of local protocols may be harder to justify. "Gateways" are required for both of these interconnection strategies.

Gateways account for variations of bandwidth, delay, error characteristics, topologies, and differing protocols between connected networks. They may have to make media and protocol translations between the networks. Media translations account for differences in the physical media employed by each network and the data link protocol and access method, and are normally straightforward. Protocol translation of higher-layer protocols is usually more difficult because these protocols provide richer and more specialized semantics, and it is not always clear how the functionality of one protocol maps into another

[BOGD80]. This is one of the reasons why ICST is defining standard higher layer protocols [BLAR80].

It is important that differentiation be made between names, addresses, and routes. A "name" tells what an object is; an "address" tells where it is; and a "route" tells how to get there [CERV78]. The network layer is concerned with addresses and routes. Names are mapped to logical addresses at the transport or session layer. The logical to physical address mapping is normally done at the network layer. The logical addresses are then associated with a higher layer protocol interface that may employ a name to describe the process being supported or the entities involved. The logical address mapping allows for the servicing of multiple entities, both in terms of the higher layer protocol interfaces and in terms of multiple logical destinations being mapped to the same physical address.

The name of an object can be any arbitrary string of characters that uniquely identifies that object. To maintain the uniqueness of names in local networks, separation of name sets must be made (e.g. all the names on one computer may be considered a set), providing a logical address. On the local network addresses are either link (on local network) or network (internet) addresses. For non-interconnected addresses these addresses would be identical. These sets of names can then be associated with a common name, where the common name could for example be the computer or network name [POUL78] [CERV78].

The format of the network packet will vary with the amount and complexity of information that the network layer provides. Elements usually included in the packet header are destination address, source address, control field, and data field.

The size of the packet is a formatting issue that affects the passage of a packet from one network to another. It is the responsibility of the internetwork gateway to fragment packets that are too large and reassemble them if they are being passed on to another network. Reassembly of fragmented packets is either handled at the exit gateway of the local network or the fragments are left for reassembly at the destination. The maximum packet size must be optimized for the network as this will affect network performance in terms of total information transferred, network overhead, efficiency, and delay.

4. Transport Layer

Transport services provide reliable end-to-end (host-to-host) transporting of messages from one transport entity to another. The transport protocol takes the variable length transport service data unit (TSDU) from its user, divides it into fixed size transport protocol data units (TPDU); and controls the passing of these data units to the peer transport protocol. The transported protocol data unit is capable of traversing any arbitrary topological configuration of interconnected networks. The Transport services cover five basic types: connection-oriented, connectionless, broadcast, multicast, and expedited. Present voluntary standards protocol developments do not include broadcast and multicast [BURJ81a].

These various services are provided to optimize the use of the network resources and for support of different applications. Each of these services will be discussed in terms of the subset of transport functions they provide. The quality of service parameters will be delineated and then the overall set of transport functions will be defined.

The quality of service can provide variations on:

1. Throughput - the amount of actual information transferred,

2. Transit Delay - the time it takes a PDU to get from one endpoint to another,

3. Residual Errors - the reliability of PDU delivery,

4. Service Availability - is the service always available and ready when requested,

5. Delivery Notification - PDUs can be individually acknowledged, acknowledged in blocks, or not acknowledged at all,

6. Sequencing - delivery of the PDUs of a SDU in the same sequence in which they are sent,

7. Flow Control - control the flow of multiple PDUs from endpoint-to-endpoint.

8. Expedited Delivery - the delivery of priority PDUs or control information before other pending PDUs.

9. Priority - establish the priority of PDUs processed by transport services,

10. Security - provide encryption and decryption, support secure routing, and verify receiver's security level,

11. Accounting - maintain record of SDU transmissions on a PDU basis,

12. Access Control - authorize the use of a connection based on security or accounting constraints.

Many of the functions of the transport layer can be associated with quality of service. While all service types have distinct qualities of service, there can be options provided within each that make these services overlap.

A connection service establishes, maintains, and terminates a connection that represents a two way simultaneous data path. A connection service can support multiple connections between transport layer endpoints The connection service maps transport addresses to network addresses and multiplexes transport connections onto network connections. This service is differentiated from other services in that the transport connection must normally be established before data can be transferred. Connection services are most useful for applications that require high throughput delivery. These high throughput communications between endpoints are normally sustained for long intervals or transfer large numbers of TSDUs. Connection services can provide the following functions:

- o End-to-end error recovery,
- o End-to-end sequence control,
- o End-to-end error detection and quality monitoring,
- o End-to-end segmenting and blocking,
- o End-to-end flow control on individual connections, and between transport and session layer, and
- o Supervisory functions.

A connectionless service provides for the simple end-to-end transfer of a single TSDU. A single TSDU can be composed of multiple PDUs, but the transaction is delivered as an atomic unit to the receiving application. A connectionless service does not require prior establishment of a connection between the transport layer endpoints. The connectionless service can provide either reliable or unreliable message delivery. Connectionless services are most appropriate in applications that require low delay. These applications are normally lower volume, bursty information transfers. Additional functions such as guaranteed in-sequence delivery and flow control of TSDU fragments (TPDUs) can be obtained with the basic connectionless services. Each of these additional functions adds to the complexity of the basic protocol and moves it closer to a connection service. Datagrams use fewer commands to transfer data from end-to-end, but require that more overhead (information about TSDU size, location, timeouts, source and destination port addresses, etc.) be carried in each command.

Broadcast services provide for the transfer of a message to all nodes on the network. Most current implementations of broadcast services are based on unreliable datagrams. The difficulty of providing reliable broadcast services is a function of the media access strategy and topologies (e.g. contention bus - difficult, token rings - easier, central polling - easiest).

Multicast services provide for the transfer of a message to a specified list of nodes, where the user makes only one request for service. Implementation of multicast services requires network layer datagram service. Connection-oriented multicast service is still a research issue.

Expedited service provides an additional means of information exchange on a transport connection. Expedited transport service data units are subject to their own set of transport-service and flow control characteristics. The maximum size of an expedited TSDU is limited. Expedited service bypasses the flow control of normal TPDUs.

5. Session Layer

A session defines a period of activity between two communicating presentation layer entities. Session control provides binding and unbinding of these two communicating entities and controls the dialogue between them for exchange, synchronization, and delimiting of data. A session exists until it is released in an orderly fashion by one of the dialogists. The session layer maintains the state of the dialogue between presentation entities [Burrj81b].

The establishment of a session includes the establishment of at least one transport connection for which type and grade of transport service and security authentication can be performed. The user can request and negotiate the individual quality characteristics of the session including:

1. Configuration of transport connections into the session (e.g. provide interrupt service),

2. Type and grade of transport service to be used,

3. Maximum size of the quarantine unit [BOLB80] [BURJ81b], where a quarantine unit is an integral number of data units that are not made available to the presentation-entity until release by the sending presentation-entity,

4. Dialogue type and interaction unit.

The session service establishes and releases the session, defines the context and identity of the session, performs session recovery, and manages the dialogue by delimiting and "pacing" of data exchanges. A single operation can also be provided to set up, use, and release a session for support of transaction processing.

Context management supports the negotiation of a high-layer-protocol (or multiple protocols) to be used between two cooperating presentation entities. Context management selects which protocol is to be used, switches from one protocol to another, and authenticates the context of a session. The authentication mechanism can be used for security and integrity checks for the session. The session layer translates the name of a process communication into a logical address. This function is performed by a name server. Context management is a subject of further ICST research.

Data exchanges can be grouped into four different functional units:

1. A data unit that delimits the exchange of data units between the user of the session and session entities. The maximum size of this unit can be negotiated as a quality of service parameter between presentation entities.

2. A quarantine unit that delimits a set of data units. A quarantine unit is only usable when it is completely received and assembled by the receiving session entity for delivery to the presentation layer.

3. An interaction unit that determines whose turn it is to communicate between two presentation entities.
4. Two data integrity units for use in recovery (roll back) procedures and resource commitment.

Interaction of presentation entities can be specified as a quality of service parameter for two-way simultaneous, two-way alternate, and one-way (one side always sending, the other side always receiving) interactions. Expedited data delivery can also be specified for sessions as for transport services.

6. Presentation Layer

The presentation layer provides management of formats and performance of transformations to the applications layer. These presentation capabilities are provided as a set of services that may be selected by the application layer. These services enable an application to interpret the meaning of data exchanged, and manage the entry, display, and control of this structured data. The mechanism used by the presentation layer to present data to the application layer is the presentation-image. The presentation-image is the data structure and actions required to make this presentation to the application layer. ICST documents describing the features and service specification of a presentation layer protocol (the Data Presentation Protocol - DPP) are presented in [CLOS80a] and [CLOS80b].

There are three phases defined for the presentation layer: the presentation-image control phase, the data transfer phase, and the presentation termination phase. Each of these phases can be executed within the context of a session.

In the presentation-Image control phase the selection or negotiation of options in an image is performed. These options may cover quality of service that indicate the acceptability, cost, or optimization value of a structure or action.

The functions of the data transfer phase are the actual presentation and transformation of the presentation-image. Control mechanisms for accessing the data structures for security would be implemented in this phase. Also special data transformations or manipulations (e.g. encryption, compression) would be done in the data transfer phase.

The protocol being used for presentation to the application layer is released in the presentation termination phase. This phase is used as a transition into the session context management phase to new presentation-images for data transfer or session termination.

Specific presentation types would provide character representation independence, data definition, command formats, conversion algorithms, etc. Examples of these presentation types and the services provided are:

1. Virtual Terminal Presentation
    1. selection of terminal class and type
    2. negotiation of profile,
    3. data and command transfer, and
    4. forms management.

2. Virtual File Presentation
    1. formatting of virtual file storage commands,
    2. communication of file data and commands, and
    3. code conversion.

3. Job Transfer and Manipulation Presentation

    1. control of record structures and devices,
    2. control characters,
    3. command formatting, and
    4. data formatting [ISO82].

7. Application Layer

Protocols required at this layer provide service-oriented functions in support of internal and external services specific to the network application. These services include specialized protocols for file transfers, virtual terminals, network monitoring and statistics, etc. These specialized protocols may dictate the design of the protocol architecture or at least influence the service types provided at each layer of the architecture. A specific example of voice application protocols and their influence on the protocol architecture is discussed below.

These discussions of voice applications are based on the DARPA's Experimental Integrated Switched Network (EISN) [MIT79] [FORJ79] [COHD78]. The EISN voice protocol architecture can be viewed as a four-layered structure consisting of:

1. Voice application protocols for conferencing and voice files,

2. A Network Voice Protocol (NVP) through which voice connections are established and managed,

3. A Stream Protocol (SP) providing basic real-time internet transport services for the voice connections, and

4. Lower layer network-specific protocols providing intranet routing.

These voice-oriented protocols vary from the standard layer protocols in routing and internetting, reliability, session control, security, flow control and monitoring requirements.

Because SP uses connections, gateways must maintain state information for each connection. This state information includes mapping tables to translate between SP connection identifiers and the network address of the next SP gateway handling the connection.

Reliability considerations are very different in a voice architecture compared to those for data transfer. Absolute bit reliability within individual packets is not necessary, nor is it desirable if it results in unacceptable delays, because interpacket timing is critical in the fidelity of the reconstructed speech, but individual bit errors will not significantly alter the overall reconstructed word or phrase. SP headers are protected by checksums to ensure the reliability of control information, but the data portion may contain bit errors. The SP headers contain the more important sequence and timestamp information.

Lost packets are not a problem in voice as long as packet size is small enough that the lost information can be smoothed over. It is desirable to maintain packet sequence and the relative timings of individual packets. This enhances the fidelity of the reconstructed speech. Therefore NVP includes sequence numbers and timestamps in the headers of data packets.

The concepts of session management span across both the NVP layer and the SP layer. There is a close coupling between the session management commands within one protocol and the session management within another to handle multi-user sessions (voice conferencing).

Security issues are addressed within various layers of the voice architecture. A major advantage of digital voice over analog voice is the applicability of standard digital encryption techniques. The presence of a central "conference chairperson" allows for easy access control within a voice conference. Access control for normal point-to-point connections can be handled at the source, within the SP gateways, or at the destination.

There is no reason for sophisticated end-to-end flow control within a voice protocol, since the data rate is negotiated between the two vocoders prior to connection establishment. The fixing of an internet gateway route for each connection guarantees the characteristics of the flow to the corresponding vocoder. Dynamic renegotiation of parameters during a voice connection can handle extreme flow control problems.

The network must be closely monitored if effective voice communications are to be maintained for the architecture. Delays in particular must be kept within certain tolerances. SP agents monitor network traffic load and delays in order to decide on connection grants. A connection request will include a flow specification which can be met only if current network and gateway conditions allow.

It is important that the network delay characteristics be closely monitored to ensure that the contracted level of service is being provided. Once a connection is established, NVP includes mechanisms where it can respond if network conditions are found to be changing. NVP can renegotiate vocoder parameters in response to changing network service quality.

2.3.3 Placement of Network Intelligence.

Placement of network intelligence impacts protocol design, in terms of interfaces between the protocol layers and the functions that are to be performed within each layer. In the ISO Reference Model [ISOR82], higher layer protocols build upon the functions provided by lower layer protocols. This layering partitions functions cleanly and simplifies protocol interfaces. The protocol layering can also simplify program development, program debugging, and operational trouble shooting. The decision as to where and how the protocol is implemented determines the allocation of intelligence in the local area network.

There are basically four approaches to providing equivalent functionality in support of interprocess communication in an environment of multiple hosts interconnected by a multiaccess communication medium. These vary in the amount of functionality which is provided by the network interface which is in turn directly related to the processing power and buffering capability of the interface device. The scale is from little processing power and no buffering capability to minicomputer processing power and storage capability.

1. Host performs all functions of a network node and drives the communication medium directly through a device driver or channel interface. This sort of approach is frequently used for close coupling of pairs of processors.

2. Host is interfaced to the communication medium by a network interface which provides the host with insulation. The host can drive the interface with commands to send and receive messages of predefined format and variable size to variable destinations. The host is responsible for formatting of messages to suit the network interface. The interface is either capable of buffering one or less messages on behalf of the host. The interface implements a physical interface to the medium and a Link control discipline.

3. Network interface adds end-to-end reliability, capability and sufficient buffering to be able to implement the buffer multiplexing necessary for the integrity of end-to-end data delivery. Host is fully insulated from the communication medium, its control and management.

4. The interface unit is large and powerful enough to contain all the protocol layers necessary for network interconnection, and designed to work in conjunction with a simple host to front-end protocol to enable user access through one or more special purpose devices. Thus an intelligent network front-end which is now very close to the user access layer.

Implementations of network protocols resident within host computers usually exist as separate processes, requiring services from the host operating system and file system. Protocols that require host specific information are large and complex, and must reside in the host.

Network front-ends are usually implemented via minicomputers and/or microcomputers, where host computers treat them as specialized devices that perform the majority of network control processing.

When making the choice between layers of protocols to be processed on host computers versus network front-ends, the following are considerationed:

1. Host-resident protocol processing can greatly decrease host efficiency, but allows for development of specialized network software by personnel already trained on host.

2. Network front-ends offload host from network processing and are usually available off-the-shelf, but specialized development must be additionally purchased.

3. Network front-ends may require the modification of existing host communication software.

2.3.4 User/Network Interfaces. Local networks require data communications support for several kinds of user or resource nodes. These include:

1. Character at a time terminals,

2. computer-based hosts,

3. Intelligent terminals, and

4. special purpose resources, such as data base and central file system machines.

To be truly effective, the local network must provide an integrated, high performance mechanism for the interconnection of this diverse set of heterogeneous resources.

Interface protocols make this happen by:

1. Using interface standards wherever possible in order to exploit vendor provided access methods for new nodes,

2. Providing the functional capability to support the range of expected traffic types,

3. Providing satisfactory performance in terms of bandwidth and delay, and

4. A growth path permitting simple attachment of nodes with minimal requirements, while providing room for expansion both in terms of performance and function.

There are various methods of interfacing user equipment to a local network. These generally fall into two classes: serial interfaces and parallel interfaces. The interface technique used impacts the cost and performance of a local network, and particularly affects the amount of user-developed software required to attach to the network. This section discusses interface classes and the tradeoffs involved.

Serial interfaces are implemented via either asynchronous or synchronous techniques, with the following categories:

1. Asynchronous, international standard (e.g. X.28).

2. Synchronous, manufacturer standard (e.g. peripheral emulation).

3. Synchronous, international standard (e.g. FIPS X.25).

Parallel interfaces are often required for very high speed applications, and may be implemented in order to provide:

1. Peripheral emulation, (IEEE 488 standard).

2. Special purpose parallel interface.

Tradeoffs between serial and parallel interfaces involve examination of the following areas:

1. Device data rate - standard serial interfaces are currently limited to 56 Kbs, while parallel interfaces can be matched to the maximum speed of the attached devices.

2. Compatibility with interface standards - vendors are providing standard interfaces and the associated protocol software that reduces the diversity of interfaces required.

3. Complexity/cost of interface implementation - parallel interfaces are more complex and therefore more costly than serial interfaces. This is due in part to the number of lines involved and the controlling and synchronization of these lines.

2.4 Transmission Media

The transmission medium is the physical connection between network transmitters (sources) and receivers (destinations), bridging the distance between them. It may be a pair of wires, coaxial cable, radio waves, optical fibers, or infrared transmission through the atmosphere. Transmission media used for local networking differ in raw transmission capacity (bandwidth), potential for connectivity in terms of point-to-point or broadcast capabilities, the geographic scope allowed due to attenuation (progressive decrease of signal power with increasing distance) characteristics of the medium, immunity to noise, relative cost considering the actual hardware purchase required and its installation, and suitable applications.

In this section we further describe these points of comparison, followed by discussions of the individual media types.

2.4.1 Points of Comparison. The following subsections describe the topics mentioned above, providing the basis for comparison between each of the various types of transmission media.

Information may be sent or propagated by means of either analog or digital transmission signals. Analog transmission refers to the use of continuously varying signals throughout the network. Digital transmission implies propagation of user information by means of discrete signal levels. If voice is transmitted via digital techniques, for example, the network must be capable of converting the continuously varying user voice signal into a digital pulse stream. The reverse conversion of a pulse stream back into a continuously varying signal is required to reconstruct the user voice at the output point. Local networks of fairly large geographic scope (intra-city) may employ a mixture of analog and digital transmission techniques along different portions of the path between points.

Most input signals cannot be sent directly over the transmission channel. Instead a carrier wave, whose properties are better suited to the transmission medium in question, is modified to represent the information. Modulation is the systematic alteration of a carrier wave in accordance with the modulating signal and may also include coding methods. Modulation is often necessary for ease of radiation, to reduce noise and interference, for frequency assignment, for multiplexing, and to overcome equipment limitations.

Many different modulation techniques have evolved to suit various tasks and system requirements. Detailed discussions of the modulation, encoding, and filtering techniques employed for local networks are beyond the scope of this guideline. Most local networks available today make use of one or more of the following basic methods [CARA75], depending on network requirements for noise and error reduction, and frequency assignment.

1. Baseband transmission. Transmission of digital signals using encoding techniques such as Manchester phase encoding is included in this category.

2. Pulse-code modulation (PCM) and its variations, using sampling techniques for analog-based information.

3. Amplitude, phase, and frequency modulation (AM, PM, FM) for analog signals, and frequency-shift keying (FSK) or phase-shift keying (PSK) for digital signals.

Bandwidth is the width of the signal spectrum, and is a convenient measure of signal speed useful for media comparison. Similarly, the rate at which a system can change stored energy is reflected by its usable frequency response, measured in terms of the system bandwidth. Speed is an important property of any network. Efficient network utilization requires minimization of transmission time, i.e. sending the most information in the least amount of time. Rapid information transmission is achieved by using signals that change rapidly with time. However signaling speed cannot be arbitrarily increased, for the system will eventually cease to respond to signal changes. The various transmission media are capable of differing information rates, but transmitting a large amount of information in a small amount of time requires wideband signals to represent the information and wideband systems to accommodate the signals. Bandwidth therefore emerges as a fundamental limitation.

Certain transmission media are capable of point-to-point and/or broadcast connectivity, due to the nature of their construction and implementation. Direct wire connections may be more applicable for point-to-point communications, while radio may be best suited for broadcast situations. Included under this topic are the limitations for multi-drops from a main segment of the media where applicable, with considerations for repeaters and/or amplifiers.

The maximum distance between points on a network is referred to as its geographic scope. Power is attenuated due to dissipative or radiative effects as it travels further from its source. With more power, information can be transmitted longer distances. The geographic scope is therefore affected by the media capacity for transmission power with and without repeaters/amplifiers, and by the amount of inherent attenuation.

Noise is a contamination to the network that alters the signal shape, an unintended signal perturbation. It is a broad classification that actually includes three effects: distortion, interference, and pure noise. Distortion is signal alteration due to imperfect response of the system to the desired signal itself, and disappears when the signal is turned off. Interference is contamination by extraneous signals, usually man-made, of a form similar to the desired signal. Pure noise is the random and unpredictable electric signals from natural causes, both internal and external to the transmission system. When such random variations are added to an information-bearing signal, the information may be partially masked or totally obliterated. Pure noise cannot be completely eliminated, even in theory. When we refer to noise, pure noise is intended.

Typical noise variations are quite small, on the order of microvolts. If the signal variations are substantially greater, then the noise may be all but ignored. The signal-to-noise ratio, measured in dB, is therefore sometimes large enough for noise to go unnoticed, but not always. Differing transmission media have varying amounts of immunity to noise, due to their construction and nature of operation.

In consideration of the above characteristics, each type of media are best suited to certain generic types of applications. For example, the broadcast nature of radio implies a quite different set of applications than the point-to-point nature of twisted pair.

Transmission media differ in relative cost, due to the nature of equipment required for information transmission, transportation, and reception; installation; maintenance; and life-cycle.

Information transmitted in a local network may travel through wire, cable, radio waves, or light beams. That physical channel produces certain constraints on factors important to the DP manager: the amount of information that may be transmitted in a given period of time, the quality of information when received, the size of the network, and its cost. Therefore, for each type of media general discussions of its construction and technology are given, with details concerning the areas for comparison.

TWISTED PAIR

Twisted pairs of wire are a form of transmission line, i.e. a means of conveying signals from one point to another. The two wires are arranged in a regular geometric pattern (spiral) in order to make electrical properties constant throughout the length of the line and to reduce noise. Radiation from the twisted pair can occur when the relationship between conductor separation and operating frequency reaches a certain point. Consequently definite limitations on frequency of transmission exist.

Transmission over twisted pair can be either analog or digital, using a variety of signaling approaches. Digital techniques often make use of pulse-code modulation (PCM).

Depending upon distance, signaling techniques, and quality of wire pair, information can be transmitted at several hundred kilobits/second and faster in point-to-point situations. If repeaters are spaced at sufficiently close intervals, speeds in the megabits/second range can be achieved. In PABX systems, typical transmission lines operate at 64 K bits/second maximum, with user device lines operating at 9600 bits/second. Multipoint applications are limited to fairly low speeds (~ 1200 bits/second) due to load and capacity considerations.

Twisted pair can be employed for both point-to-point and multipoint applications. When used for multipoint situations, however, average device data rates are severely restricted, dropping as the distance increases between devices due to propagation delays. Device burst rates may still be high, but a large number of devices will restrict concurrent usage for all devices. Point-to-point usage is far more common and most suitable for general local network applications.

Twisted pair can range over an area of fifteen kilometers, provided that suitable conditioning takes place throughout the length of the line. Most implementations utilizing twisted pair restrict distances to within buildings.

Energy loss is an important parameter to consider when discussing geographic scope (and network security). As distance between communicating devices increases energy loss will also increase, such that at some point the receiver will not be able to properly respond to and detect incoming information. The lost energy radiates to outside the transmission line, allowing motivated outsiders to detect and pick up that energy. There are two ways in which energy, impressed at the sending end of a transmission line, may become dissipated before reaching its destination: radiation and conductor heating. Radiation losses arise because a transmission line may act as an antenna if the conductor is an appreciable fraction of the transmitted wavelength. Such losses are difficult to estimate, being normally measured rather than calculated. Conductor heating is proportional to current and impedance, increasing with frequency as the signal is carried more toward the outside of the conductor.

Twisted pairs of wire can be purchased with a variety of properties at a variety of costs. They are available with a varying number of twists per foot, with no electrostatic shielding, with braided shields, or with solid shields. In general, good noise immunity can be achieved for noise with effective wavelengths much longer than the "twist length" of the cable. In balanced low frequency systems, noise immunity can be as high or higher than for coaxial cable. However, at frequencies above 10 to 100 kHz, coaxial cable is typically superior.

The high shunt capacitance of this cable can cause signal distortions. Quality of analog signal transmission can be improved by utilizing devices known as loading coils. These inductive components offset the shunt capacitance, which builds up as the length of the communications line increases. These loading coils tend to enhance voice transmission quality within the 0 to 4 kHz voice band but are not especially useful if the line is to be employed for wideband data transmission.

The point made is that an extremely wide variation in noise immunity and other properties for twisted pair products and for networks utilizing those products are available; therefore close examination of requirements versus product limitations should be made.

The most common contemporary network implementation utilizing twisted pair transmission lines between network nodes is the PABX (private automated branch exchange), principally used in telephony. Many PABX systems utilize digital encoding and switching, and can be modified for data switching as well. Twisted pair connections are usually from device to central switch or intermediate concentrator.

The twisted pair medium itself has been used for a variety of communication applications, but its connectivity and noise immunity limitations must be recognized. Unless properly protected, it should remain within buildings or rooms. When part of a bus network, for example, typical applications are for device to computer or network interface unit and not as the general network medium.

Twisted pair is useful in multipoint applications when very low speed, low duty cycle devices are to be interconnected. A common usage is for the connection of traffic lights to their central controller.

Twisted pair is the least expensive of the transmission media examined when looking at cost per foot of wire. However, considering the amount of conditioning that might be required and its connectivity limitations, installation costs may approach other media, such as coaxial cable.

COAXIAL CABLE

Coaxial cable is a form of transmission line very similar in concept to twisted pair, but with modified construction to provide different operating characteristics. The cable has an inner conductor (wire) with an outer conductor concentric with and completely surrounding it, which is usually grounded. Between the inner and outer conductors is a dielectric (air or solid material); the entire cable is housed by an outer casing.

Coaxial cable currently used for local networking are generally classified in two ways according to the modulation techniques employed: baseband and broadband. There is a slight physical difference between the two types of cable: baseband coax usually consists of a carrier wire surrounded by a woven mesh of copper and is typically 3/8" in diameter, while broadband has a sleeve of extruded aluminum and is slightly wider.

Baseband coaxial cable are usually of 50 ohm grades, capable of sending a single signal using digital techniques. Connection to the cable is often through a nondestructive tap with a passive transceiver. Highly reliable repeaters, amplifiers, and taps are available. Considering that transmission is without modulation (relatively low transmission frequency), construction issues of baseband coaxial cable (e.g. relationship of conductor to transmitted wavelengths) may present security problems, in that the cable may act as an antenna allowing eavesdroppers to tap into the line with pickup coils appropriately placed.

Broadband media concepts are implemented primarily with off the shelf CATV (Community Antenna TeleVision) hardware, generally using 75 ohm cable for dual and mid-split systems. Low cost signal splitters and taps achieve branching of cables, and commercial repeaters (line amplifiers) ensure adequate signal levels through the system. Mean time between failures for these CATV devices has been established at 400,000 hours or better in large installations. Fully redundant repeaters are also available, and status monitoring equipment for these units has recently been introduced. All equipment is mass produced and extremely reliable due to heavy use by the CATV industry.

Baseband transmission implies no modulation of the digital signals transmitted; transceivers drive data onto the cable using a variety of coding techniques, such as Manchester phase encoding. Information is transmitted in bit serial fashion; one signal occupies the cable bandwidth.

Networks using broadband coaxial cable often employ frequency and phase modulation techniques, transmitting analog signals along the cable. In its most simple form, a single broadband cable may be characterized as a two-way radio frequency medium; for subsplit systems with a forward bandwidth of 140 MHz and return bandwidth of 105 MHz. Signals are transmitted on a return spectrum channel and retransmitted at a higher frequency on a related forward channel. All points attached to the broadband coaxial cable can receive the retransmitted signal.

Bandwidth differs significantly according to the mode of transmission. Existing baseband coaxial cable network implementations are limited to a single signal, with backbone cable rates between 3 and 10 M bits/second. Such limitations are not necessarily a result of cable technical properties but of total network cost, with manufacturers making the resulting tradeoffs. High speed multipoint implementations do exist (~50 M bits/second) with specialized high-cost hardware.

Broadband coaxial cable has a capacity that is midrange between fiber optics and baseband cable. For each trunk (main cable segment) an ultimate of approximately 150 M bits/second of full-duplex transmission path is available, for a total of 300 M bits/second. All channel assignments are done by frequency-division multiplexing, where a channel may be as wide as the full capacity of the cable. Considering that network manufacturers desire to build systems that can share cable with existing operations such as CATV (6 MHz), that bandwidth may be cut accordingly.

Coaxial cable is applicable to point-to-point and in particular to multipoint topologies. Conventional topologies may be implemented (star, ring, bus, mesh), and the complexity can range from simple to sophisticated. Considering the variety of access and multiplexing techniques available, variations of the bus topology are particularly applicable. Systems can very often be implemented using only one cable.

Multidrop capabilities of a single cable segment for the two types of cable are highly dependent upon applications and desired data rates. Baseband multidrops are on the order of 100, with broadband supporting several thousand; the basic difference stems from whether single or multiple signals are able to transmit simultaneously. Repeaters and/or amplifiers must be placed approximately every 600 to 1600 meters to regenerate signal shape or amplify signal power to original levels, respectively.

Depending on tolerable delay, load, and implementation, maximum distances in typical baseband coaxial networks are limited to one to three kilometers; broadband networks can span areas of ten kilometers or more (fifty is a practical upper limit).

The basic difference between the two distances lies in technical considerations based upon information transmission mechanism (analog or digital), modulation techniques, transmission frequency, and attenuation properties of the two types of cable. The types of electromagnetic noise usually encountered in industrial and urban areas are of relatively low frequency. Therefore information transmitted at baseband in digital form (square waves) may be highly susceptible to that noise. Repeaters placed along the transmission path may employ filters to remove noise, but limitations on distance do exist. Analog information modulated on a carrier is less susceptible to the types of noise frequently encountered. Modulation techniques that reduce noise effects coupled with higher frequency transmission (low frequency noise would not hamper reception) allow a wider geographic scope for broadband.

Inherent attenuation of the types of cable utilized must also be considered.

Immunity to noise for coaxial cable networks is highly dependent upon their application and implementation, as described in previous sections. Typically baseband systems have an immunity of 50-60 dB in isolation, while broadband has a similar figure of 85-100 dB.

Coaxial cable networks are particularly applicable to sub/tree architectures, such as might be useful for office automation, laboratory, and process control environments. Devices requiring network communication may be scattered throughout an industrial complex, tapping into appropriately placed cable segments. Depending upon implementations, devices may communicate at a variety of data rates, and hence may range from low speed interactive terminal links at 300 bits/second to high speed computer to computer links (2-5 M bits/second). Broadband coax has large capacity supported by high multidrop capability, applicable to data distribution networking. In addition the cable can be shared with other networks or systems.

Per-foot cost of the coaxial cable itself is not high, although more expensive than twisted pair but currently less than fiber optics. Depending on the application, however, total installation cost may actually be lower than that for twisted pair, considering the multipoint capabilities of coax versus the latter's point-to-point limitation.

Due to their differing properties and capacities, the cost of broadband coax is approximately 1 1/2 times that of baseband. Office automation environments tend to require fairly low cost services, making coaxial cable quite viable.

FIBER OPTICS

Optical fiber transmission is implemented by transmitting a signal-encoded beam of light through an optical cable, consisting of a group of discrete optical fibers that each transmit a light signal from one end of the cable to the other. Transmission takes place within the infrared frequency range, 10**14 to 10**15 Hz.

A single optical fiber has a center core of a glass or plastic material with a high index of refraction, surrounded by a cladding layer of a material with a slightly lower index; the relative difference in the index of refraction between the core and the cladding is approximately one percent. The transmission of light is based upon total internal reflection of the light as it travels along the core of the fiber. The cladding layer thus isolates the fibers and prevents cross talk between adjacent fibers [SCHM80].

Three distinct fiber types are possible. The first type is known as single-mode step index fiber, where the core diameter is extremely small, typically used for transmitting a single mode of linearly polarized electromagnetic radiation (monomode transmission). The second type is a multi-mode step-index fiber which allows a larger number of propagation modes, and is bandwidth limited for long distances. The third type is the multi-mode graded index fiber, where the index of refraction falls off gradually from the center of the fiber toward the outside; this fiber has been demonstrated to have the highest transmission rate over the greatest distance of the three, but also currently with the highest cost. Such costs are decreasing as manufacturing increases and improves [LITI80].

Fiber optic cables are capable of supporting several optical fibers. Some cables include a steel stablizing central member, but this type of cable should be avoided in applications where computer security is of importance. By possibly acting as an antenna to the computer, the metal cable mitigates one of the key advantages of fiber optics, i.e. that eavesdroppers can no longer tap the line by placing a pickup coil nearby.

Semiconductor lasers and light emitting diodes (LEDs) are the dominant sources for light-wave transmission [BERA80]. Lasers are ideal because they couple power more efficiently into the fibers and because their narrower spectral width reduces the effect of the intrinsic chromatic dispersion of the fibers. LEDs are adequate for data links and less costly. The environmental temperature range encountered by electronics used in networks is quite large, accommodated easily by LEDs and difficult for lasers. Lasers have an emission threshold intrinsically temperature sensitive, leading to a need for circuitry to automatically adjust the driving current with temperature. Currently available lasers also have a significantly shorter operating life than do LEDs. Both LEDs and lasers will be needed for different applications in the future.

As described below, optical fibers are unaffected by electromagnetic interference and noise, cross-talk, and electrical shorts; multiplexing of light frequencies (colors) is presently not practical. Many of the reasons for modulation are therefore not present, and consequently fiber optics is most suitable for baseband (no modulation) transmission in digital form for point-to-point applications.

Under laboratory and other experimental conditions, data rates as high as a few gigabits per second over a single glass fiber have been demonstrated. Practically, rates between 1 and 50 M bits/second over a distance of ten kilometers are easily achievable. Using mass- produced transmission and reception equipment it is possible to operate at 140 M bits/second over a distance of six to eight kilometers without intermediate repeaters.

Optical fibers and cables can be connected via both point-to-point and broadcast (multidrop) methods. Present network use of fiber optics is mainly limited to point-to-point communications however, but experimentally and expensively the medium can operate in a multidrop environment using optical couplers. Couplers are used to directionally combine optical energy from two or more waveguides into one waveguide or to split energy from one waveguide into two or more. Research has been conducted in the area of multiplexing techniques for fiber optics, where different light frequencies (i.e. colors) are multiplexed and demultiplexed, but these optical multiplexers are not presently practical. Passive multidrops are currently limited to approximately sixteen nodes. Multidrop network implementations must also recognize the unidirectional characteristics of fiber optics. Two-way communications, such as a bus, must use two fibers; in ring-type architectures a single fiber is sufficient.

While not presently practical, a single segment of fiber optics could support many more drops (an order of magnitude) than coaxial cable, due to lower power loss at each drop, lower attenuation characteristics, and greater bandwidth potential.

The composition of the fiber itself determines the transmission attenuation, caused either by scattering or absorption by trace elements present in the core. The manufacturing process of such fibers, using ultrafine and ultrapure compounds, is therefore critical in producing the fiber with the most desirable characteristics, or optimum tradeoff between cost and attenuation.

Fiber optics exhibit losses in the two to ten dB/kilometer range. Losses as low as 0.5 dB/km have been demonstrated under laboratory conditions. Using repeaters and/or amplifiers where required to regenerate digital signals and/or increase power to source levels, fiber optic networks could span a distance of 50 kilometers or more. Present technology in production allows transmission over a distance of six to eight kilometers without intermediate repeaters at high speeds.

Optical fibers are unaffected by electromagnetic interference and noise. Since light does not radiate through the cable, cross talk (unwanted coupling of signals from one channel to another) does not occur. The danger of electrical shorts between conductors is gone, and should the fiber optic cable suddenly be immersed in water, the signals will continue to propagate unhampered.

In consideration of current cost of creating multidrops from the cable, fiber optics is presently most suitable for high-speed point-to-point applications for local networks. Examples include:

4. Computer-to-computer high speed link.

5. Connection between a terminal and a processor over kilometer range distances.

6. Link between buildings of an industrial complex.

7. Communications path between complexes at opposite ends of a city.

Of the present local network transmission media technologies available, fiber optics implementations are more expensive than twisted pair and coaxial cable in terms of cost per foot of cable and required equipment (transmitters, receivers, connectors). The costs are declining as engineering and manufacturing techniques improve; consequently it may become a viable alternative for all local network topologies in the future. Current expense of optical couplers, taps and multiplexing techniques and devices prohibits extensive use for multidrop applications.

RADIO

Local network utilization of radio as a transmission medium revolves around packet radio technology [KAHR78]. The advantages of multiple access and broadcast radio channels for information distribution and computer communications have been established, and several experimental digital radio networks are in operation. Packet-switched communications techniques are employed, and are particularly important for computer communications in the ground mobile network environment.

Many technical problems exist; allocation of radio channels is an example. The choice of radio channels for any communication system is a complex task, requiring tradeoffs of many factors such as desired bandwidth, area coverage, spectrum availability, potential interference and noise sources, regulatory requirements, and frequencies where spectrum crowding is less severe and the availability of bandwidth is greater. Crowded radio bands are undesirable, not only because of interference to other users but also because of interference from them.

The packet radio signaling waveform must be designed to perform well with respect to both the natural environment and the induced environment arising from both intentional and unintentional interference. Such interference includes system self-interference arising from the multiple access/random access nature of the packet radio system. Spread spectrum techniques reduce the multipath-caused limitations on signaling rate, and provide rejection of interference and the ability to coexist with other signals in the RF band.

Two major types of spread spectrum signaling techniques are well suited to packet radio applications: direct sequence pseudo-noise (PN) modulation, frequency hopped (FH) modulation, and hybrid combinations of the two [KAHR78]. Bandwidth is expanded by these techniques, but additional advantages include:

o Signal to noise ratio is improved by a factor called the processing gain.

- Separation ability of the various multipath signal components, allowing recombination with reduced signal fading over time, improved signal to noise ratio, and frequency selective fading only over small portions of the band.

- Lower electromagnetic profile since the signal is spread over a wider bandwidth and its waveform is of a pseudo-random nature.

- Strong ability of receiver to correctly receive one packet in presence of other interfering packets (capture capability).

- Users can coexist in same area of frequency band with reduced interference.

- Reduced overlap of signal with delayed components.

- Reduced multipath fading effects.

The operational characteristics of the radio frequency band have a major impact on the packet radio design. The lowest and highest frequencies that can be used for packet radio are determined primarily by consideration of bandwidth and propagation path loss. Bandwidth limitations are imposed by the operational characteristics of the radio frequency band and by network requirements. Cost-effective radio equipment is difficult to achieve if the ratio of RF bandwidth to RF center frequency is much larger than 0.3. This lower bounds the range of acceptable RF center frequencies. In practice, a center frequency well in excess of this lower bound is desirable if the received would otherwise have too wide a multipath spread. For example, if a packet radio system is to deliver 2000-bit packets through a network with delays on the order of a tenth of a second, the data rate of the system must be in the range of a few hundred kilobits per second, implying bandwidths of a few hundred KHz [KAHR78]. From an implementation viewpoint the RF center frequency should be at least a few MHz, in the lower high-frequency band (HF) extending from 3 MHz to 30 MHz. Propagation in the HF band does allow long-distance communication, but limits data rate due to multipath spreading. Line-of-sight propagation dominates in the VHF (30 MHz to 300 MHz), supporting data rates on the order of a hundred kilobits due to reduced multipath spreading.

By its nature, radio is suited for both broadcast and point-to-point connectivity. Numerous technical problems do however result from station mobility, propagation characteristics of the frequency band, natural and man-made interference, reflection, and noise.

Ground-based networks encounter the most difficult environment in terms of propagation and RF connectivity. Ground radio links are subject to severe variations in received signal strength due to local variations in terrain, man-made structures, and foliage. Reflections give rise to multiple signal paths leading to distortion and fading as the differently delayed signals interfere at a receiver. RF connectivity is difficult to predict and may abruptly change in unexpected ways as mobile terminals move about. The implementation of the packet radio system should have a self-organizing, automated network management capability which dynamically discovers RF connectivity as a function of time, providing area coverage with full connectivity.

The distance between points communicating by packet radio is highly dependent on the propagation characteristics of the employed frequency band. Propagation in the HF band can provide long distance communication due to sky wave reflections from the earth's ionosphere, but the propagation suffers from noticeable multipath spreading of the signal which limits the data rate. As the operating frequency rises to a practical upper limit of 10 GHz, absorptive losses due to the atmosphere and rain rapidly increase, reducing the resulting radio range. Since local networks have an intra-city scope of a few tens of kilometers, radio frequencies in the upper VHF (100 MHz to 300 MHz), UHF (300 MHz to 3 GHz), and lower SHF (3 GHz to 10 GHz) bands can be utilized to provide that range. Note that closely spaced relays must be utilized to provide adequate area coverage at these frequencies.

Reflections caused by local variations in terrain, man-made structures, and foliage allow multiple signal paths to the destination, imposing distortion and fading as the differently delayed signals interfere when received. Operational characteristics of the frequency band employed may cause multipath spreading, fading, and distortion, particularly at HF and lower VHF. The effect is that additional attenuation of the signal may be observed when the receiver is located at a ground point where the signal interference is destructive.

Radio is affected by both intentional (jamming) and non-intentional interference; requirements for resistance to interference strongly affects the details of modulation techniques selected and system complexity. Unintentional interference results from automobile ignitions, trains, radar, etc., and is often generated at relatively low signal strength levels. However, in urban areas the density of interference sources will be quite high, causing levels of 60 to 80 dB above the thermal noise level of the receiver. Packet radios might therefore experience bit errors in every packet received, requiring forward error correction to maintain system throughput.

Packet radio is a technology that extends the original packet switching concepts which evolved for networks of point-to-point communication lines to broadcast radio networks. Development has been greatly stimulated by the need to provide computer network access to mobile terminals and computer communications in the mobile environment. Because of its capability for dynamic allocation of the spectrum, packet radio offers a highly efficient way of using a multiple access channel, particularly with mobile subscribers and large numbers of users with bursty traffic and can also provide a degree of flexibility in rapid deployment and reconfiguration not currently possible with most fixed plant installations.

Packet radio will be essential for military and other governmental needs as terminals and computer systems become pervasive throughout essentially all aspects of their operations. Initially the needs for radio based computer communications are expected to be prevalent in training on or near the battlefield and in crisis situations. The first operational systems are most likely to be deployed for use in one of these areas where a higher relative cost of providing the advanced capability can be tolerated.

Within the civilian sector, there is also a strong need for terminal access to information in the mobile environment, but the cost of services to the user will dictate when such capabilities should be publicly provided.

From the above discussions it is clear that radio technology for local networks is still in its infancy and quite experimental. Present application might involve point-to-point or store-and-forward transmissions between buildings or across a city, where transmitting/receiving equipment may be placed in direct line-of-sight to avoid urban effects.

Considering the experimental stage of packet radio technology, complete network costs would be quite high considering that extensive custom development would be required. Radio transmitters and receivers are readily available, but total network capabilities are not presently available off-the-shelf.

INFRARED

In addition to fiber optics, information can be transmitted at infrared light frequencies through free-space (i.e. atmosphere). Light sources for the infrared transmissions include both LEDs and two types of lasers. Gas lasers have the disadvantages of higher cost, greater bulk, and high voltages, while semiconductor laser diodes and LEDs have low cost, small size, high efficiencies, a range of wavelengths, and long lifetimes. They also require low operating voltages and can be directly modulated, but have a high degree of divergence, wider spectral widths, and low duty cycles. A laser diode is better for longer ranges than an LED, due to its smaller divergence, larger peak powers, and smaller spectral width allowing improved background noise filtering. The LED however, is capable of the higher duty cycles necessary for high data rates, longer lifetimes, and lower cost.

In addition to generating the light via the sources mentioned, it obviously must also be detected at the receiving end. Direct detection is most practical for severe operating conditions, in consideration of atmospheric conditions which perturb the phase and directionality of the wavefront. The most common detectors are semiconductor PIN photodiodes, avalanche photodiodes, and dynamic crossed field PMT, all with differing characteristics [POWC79].

In order to protect against atmospheric effects and background noise, short pulse with high peak power modulation techniques are utilized. Pulse modulation has made a wide variety of coding formats possible, using both analog and digital inputs. Digital coding formats adapt well to pulse modulation, often with significant performance advantages over analog transmission, but do require more complex circuitry and larger bandwidths.

Pulse-rate modulation can partially overcome the atmospheric scintillation effects [POWC79]. Pulse-position modulation (PPM) and differential pulse position modulation (DPPM) offer the most promising aspects, such as fewer affecting atmospheric problems than with other modulation techniques, low harmonic distortion (less than 2 percent out to 3 kHz), and low power requirements.

Most systems presently in operation use laser diodes and LEDs as sources, with semiconductor detectors. Data rates therefore fall between 10 to 100 K bits/second with 10(-6) error rates over a range of approximately 16 kilometers. Reliable data rates up to 1.5 M bits/second are possible over a 1.6 kilometer path. A number of optical communication link systems are presently marketed, with data rates and distance closely tied.

Present free-space infrared communications technology is limited to point-to-point links (star, ring topologies), although recent advances make use of the infrared properties that allow easy confinement of transmission within a desired area of reception (i.e. a room) in a true broadcast mode. A central infrared station installed in a ceiling is able to transmit and receive to/from specific points within the room, using diffused nondirectional radiation. Signals are reflected in all directions, filling the room and eliminating the need for line-of-sight transmissions.

The atmosphere imposes severe restrictions on the range and error rates for infrared communications, due to absorption, scattering, turbulence, refractive index variations, and scintillation. This effectively limits optical communication through the atmosphere to ranges of less than 160 kilometers, and to less than 1.6 kilometers for the large bandwidth promised by light-wave communications. The average range is 1.6 kilometers, with the practical maximum range between 16 and 32 kilometers.

Infrared light transmissions are unaffected by electromagnetic interference and noise. Atmospheric scintillation modulates the intensity of the beams at 1 to 200 Hz rates, and the beam wander and spreading caused by refractive index changes along the transmission path will often cause sudden deep fading. Placing analog information on an FM subcarrier modulates the intensity of the laser beam, lessening the importance of nonlinear operation of the system subcomponents, changes in the laser amplitude, and atmospheric effects.

Free-space infrared communications applications have been primarily limited to point-to-point line of sight, such as between closely located buildings. The technology is still in its infancy, but broadening of the applicable areas is beginning to take place. Manufacturers have created terminal networks within a room, where an infrared source in the ceiling, for example, transmits/receives to/from terminals in that room using diffuse radiation (see above). Advances are expected to continue, and the medium may become viable for general network application in the future.

As might be expected, free-space infrared communications devices are quite expensive. Costs will drop as technological and manufacturing improvements are made, but that will take time. Note that no general network capabilities using such a medium are available, meaning that extensive custom development would be required to create and install such a network at the present time.

2.5 Network Summary

Section Two has examined features of contemporary local networks according to suitable applications and services available, types of topology, protocol architecture, and transmission media. Analysis has been oriented toward the limitations and capabilities of each feature and the tradeoffs that must be made to suit the intended application. Proper design is the key to a network that meets the requirements and maintains a long life-cycle.

Distinctions between local network applications revolve around four basic issues: the type of information that is transmitted (data, voice), the types of equipment that require communication (terminals, computers), advanced services that may currently be required by users or for which applications may evolve, and any specialized environments.

Network topology graphically defines the interconnections between nodes, and is of four basic types: star, ring, bus, mesh. No one type can be considered the "best", as they differ in areas of reliability, expandability, and performance and offer particular advantages in varied applications. Access and management strategies effect the characteristics.

The protocol architecture design provides the basic services to users and manages information transmitted over the media. Based upon the International Standards Organization (ISO) Reference Model, the architecture is divided into seven layers. Each layer has distinct functions to perform, services to provide to the upper layers, and requirements expected of the lower layers. If the protocol architecture used maps to the ISO Reference Model then the implementation of new protocols or quality of service has a minimal impact on the architecture as a whole.

Transmission media differ in the following areas: transmission technology, available bandwidth, connectivity potential, geographic scope, noise immunity, suitable applications, and relative cost. Communications technology is expanding and improving, such that twisted pair, radio, and coaxial cable may be replaced by fiber optics and free-space infrared within the next decade in both point-to-point and broadcast situations, considering their high bandwidth and electromagnetic noise immunity.

3. DETERMINING LOCAL NETWORK REQUIREMENTS

This section describes a method for determining the requirements of a local area network. They are organized in five categories: services, traffic characteristics, reliability, growth, and maintenance.

Services:

Network services define the functions performed by the network that are most visible to its users and management personnel, including communication capabilities at various levels of complexity.

Traffic:

The characteristics and volume of network traffic not only define the communication paths required between local and remote network nodes, but also their associated performance requirements.

Reliability:

The network must be available to its users a very high percentage of the time, and there must be a long period of time between component and total network failure (mean time between failures, MTBF) and a minimal period before repairs are completed (mean time to repair, MTTR). Such considerations apply to the network as a global entity, and to its links, nodes, interface units, and control units, grouped according to usage/location.

Growth:

Growth applies to all requirement categories, and defines the planned dynamics of those requirements (additions and changes) until a given future date.

Maintenance:

Once a network has been installed and accepted, maintenance at various levels may be required to ensure its availability for the future.

Each of the subsections that follow presents a description of a generic group of requirements, defining terms and providing examples of applicable alternatives. Following that description is a general discussion of the steps managers and staff might take in determining the requirements most suitable for their network installation.

3.1 Required Services

An important step in the local network procurement process is the accurate determination of the installation's networking requirements. This section describes how the procurer (or designer) determines the set of network services required by the installation.

Network services refer to the functions performed by the network as seen by a protocol layer, an application program, or user at a terminal. They range from unreliable transmission of individual packets to applications used by the agency in fulfillment of its mission. The installation's administration decides which communications functions are provided by the network and which are implemented by in-house personnel.

3.1.1 Example Network Applications and Services.

Services provided by the local network support the communications anticipated by the target installation. The network procurer generally knows the applications the network must support. From this knowledge a set of required network services are derived. Applications typical of the office environment [BIBK81b] include data base access, centralized file access, message processing and electronic mail, bulk raster traffic, graphics, real-time raster traffic, and teleconferencing.

The required set of services architectually fit within the structure of the International Standards Organization (ISO) Reference Model for Open Systems Interconnection ([ISOR80], as examined in Section 2.3). The services described below directly affect user capabilities, and can be categorized as follows:

Supported Equipment:

Equipment types range from low-cost/low-performance interactive terminals to higher-cost/high-performance computers. Within that range a variety of peripherals exist with varying degrees of intelligence and specialized functionality.

Interfacing Requirements:

Each piece of equipment requires specific interface protocols for communication with other devices. These protocols reside at the application layer of the ISO Reference Model and may use presentation layer services, such as those being developed under the NBS Program in Computer Network Protocol Standards [BLAR80].

Transmitted Information:

Information transmitted throughout the network may be one or more of the following, with differing traffic characteristics: data, voice, video, imagery.

Basic Network Service:

Basic network services allow transmission and reception of information from one point to another, where varying modes of communication may be required. Such services reside within the network, transport, and session layers of the ISO Reference Model.

Higher-Layer Functions:

The network may be required to provide application services at higher-layers than transport and session based upon the characteristics of the information processed by specialized network nodes.

Security and Privacy:

Isolation between multiple user groups and access control may be required to enforce data integrity and protection. Depending upon environment and exposure, security from outside accidental and purposeful threats may also be necessary. Security and privacy considerations must be included at all protocol layers of the ISO Reference Model, with encryption capabilities typically included at the presentation layer.

Network Control:

Network management and control mechanisms may be required for configuration control, maintenance, and accounting purposes with visibility into the operations and performance of the network as a whole. Statistics kept are accessible by network management personnel only, where visibility should not transcend into the content of information transmitted.

Remote Communications:

The ability to communicate with other users and processes on other local networks or on long-haul networks, via the network layer of the ISO Reference Model, may be necessary in order to perform agency functions. Each gateway to an outside network may impose specific interfacing requirements.

In addition to these categories of services, the network must satisfy certain physical constraints. Very rarely are special rooms, buildings, and geographic areas constructed according to network requirements; most often the network must fit into an existing set of structures with specific climatic conditions and equipment configurations. Since these physical requirements are as fundamental as the determination of required services, we will discuss them as well.

The following subsections describe each of these areas in detail, with suggestions for analysis that might be performed in order to make decisions concerning each subject.

3.1.2 Physical Network Environment. Climate and configuration are two requirements to be considered when planning the physical installation of the local network. Determination of these requirements is straightforward. Site selection and geographic scope is made considering current agency operations and planned development.

The network may be required to fit into a rigid environment in which a system is currently in operation and hence cannot tolerate even minimal impact; at the other extreme the configuration may allow new building construction to take network requirements into consideration. Actual requirements fall somewhere within these limits. Basic issues that must be considered include: the number and locations of the buildings that are to house the network, the placement of equipment to be supported by the network within those buildings, and allowable space taken up by the network and its interfaces.

In designing a network configuration that does not yet exist, analysis is made of the communication paths matching functions that the system as a whole performs [BOOA80]; minimal distance between closely functioning entities is the goal. Such analysis are usually performed and decisions made at the time procurement of a local network was determined to be the appropriate solution.

The number and locations of the buildings in which the network is housed are determined according to expected functional communication paths. If groups of equipment and/or personnel in buildings at opposite ends of a city require frequent network communication with optimum response, then the local network must be capable of supporting all such buildings at intra-city distances. If functional groups of equipment are more than approximately 1500 meters apart, but require infrequent communication where completion time is not a primary consideration, then it may be sufficient to specify separate local networks with a gateway connection to a low-speed intra-city link in each of the two buildings. Simple analysis such as described, performed on an iterative basis as plans become more firm and detailed, help determine the proper building interconnectivity requirements for a multi-building complex.

Similar analysis to that for building interconnectivity is performed for equipment locations within a building. Equipment should be colocated with personnel that interact; similarly closely communicating pieces of equipment should be colocated.

The space (height, width and depth) requirements of equipment to be interconnected to the network should already be known; remaining space available for the equipment which implements the network, in terms of network interconnection units, centralized control units, repeaters, amplifiers, taps/transceivers, and cables should be considered as well. Space requirements include:

- Reasonable room space available for network interconnection units colocated with supported equipment;

- Tolerable size and weight of each individual unit, considering expected frequency and desired ease of movement.

- Available wall and/or floor space for taps/transceivers.

- Area within or outside of walls, ceilings, and floors for cabling and repeaters/amplifiers.

- Any dedicated rooms or floor space available for network control units placed at a central location or distributed throughout the complex.

The network must be capable of operating within the climatic conditions of the buildings in which it is to be housed; in addition cables and repeaters/amplifiers between buildings may encounter direct exposure to weather, interference, and noise. Network equipment to be located indoors must be able to operate in an enclosed, normally heated building with normal exposure to dust, humidity, and temperature changes. Specific indoor climatic conditions include:

1. Conditions experienced while operating: temperature, relative humidity, and barometric pressure.

2. Non-operating conditions (temperature, relative humidity, barometric pressure) for long-term storage without damage or deterioration, and with no special preparation required before storage.

3. Normal shock and vibration as experienced during shipment, routine maintenance, and installation.

In addition, network equipment must not modify the climate in which it is to operate beyond certain extremes; it must not require more than a given amount of dedicated cooling and must not produce heat or acoustic noise levels in excess of given specifications.

Equipment may be required to operate within given electrical constraints, such as power voltage, phase, and frequency; such conditions are usually of importance when the network must be phased in with an operational system. Details concerning such design issues are dictated by the current system.

3.1.3 Supported Equipment. A fundamental service which a network may be required to provide is the support of special equipment necessary for specific applications. Types of equipment requiring support minimally include:

Terminals.

Computers.

Peripherals: printers, file storage units, tape drives, special equipment.

Terminal support is required if applications desire to communicate directly with terminals over the network, rather than only with those terminals directly attached to a specific host. Support is either through special devices allowing terminals to attach directly to the network, or through special software within hosts that allow terminals of varying characteristics to communicate with host applications. Typical of such software are the "virtual terminal protocols" which map terminal parameters into a common network-wide form [SHEC81].

Terminals fall into certain natural "classes" (e.g. asynchronous character terminals), and a network's terminal support is typically limited to only certain classes [SHEC81]. When a particular class of terminal requires network support, an explicit network requirement is made.

Support of particular computers is a more subtle issue since each computer may require a specialized interface to attach to the network. Vendors of local network products support some computers and not others. The important issue of interfaces is discussed under the next heading.

3.1.4 Interfaces. Each device included in the network requires an interface. The following attributes help distinguish interfaces: asynchronous, synchronous, serial, parallel, or custom. Generic interfaces exist as standards such as EIA RS-232C and IEEE 488.

These are the interfaces between equipment within the users domain and equipment provided by the network vendor. Since an installation may require (for perhaps non-networking reasons) that specific devices be supported, an important network service requirement is the interfaces it can export. It may therefore be necessary for the network to contain application layer protocols serving as device/process controllers, and presentation layer protocols providing data translation for device independence.

3.1.5 Hardware Interfaces. Terminals and similar devices require serial lines running an asynchronous character-at-a-time protocol. Such interfaces are nearly universal within the computer and communications industries, and by and large most network vendors provide some support for such interfaces. Standards for such interfaces are described in FIPS publications [NBSL81].

The attachment of computers to a network can be accomplished via serial interfaces, but such an attachment strategy does not take advantage of the high transfer rates possible with more advanced computer I/O architectures. Although computers tend to have a wide variety of I/O architectures, the market share of some vendors has established de facto standards which are consequently supported by network vendors. Standards for parallel high speed interfaces are described in FIPS PUBS 60 and 61 [FIPS79a] [FIPS79b].

The network procurer should catalog the specific hardware interfaces that the network must support, based on the devices which the user applications require.

Compatibility between higher-layer protocol software on each side of the hardware interface is also important. An example is provided by FIPS X.25. As with the hardware interfaces, de facto standards can emerge, and many network vendors will support such software interfaces. As vendors comply with FIPS standards that provide common software and hardware interfaces, the multiplicity of interfaces the network is required to support should decrease, and the number of vendors providing those interfaces should increase.

Thus, along with a catalog of the necessary hardware interfaces, the network procurer must establish software interfaces that the network supports. This can be based on an analysis of the attached equipment.

3.1.6 Transmitted Information. Four basic types of information can be transmitted through the network: data, voice, video, and imagery. Each impose differing implementation considerations on a local network, as described in Section 2.1.1. This section identifies the nature of the four generic types.

Data has the following typical attributes:

Message transfers are relatively short transactions, often less then 4 thousand bytes in length. They include status and control, terminal commands and results, data base access commands, human message interchange.

File transfers (stream traffic, bulk transfers): applications include data base results, archiving, inter-process file communication.

Data may or may not require transmission in real-time. The interfaces required by digital equipment to be connected to the network may make it necessary to transfer data in bit-serial or byte-, word-, or double-word parallel fashion as processed by presentation layer protocols [CLOS80a] [CLOS80b].

Protocols handling the transmission of data traffic are fairly well understood, and the ISO Reference Model for Open Systems Interconnection [ISOR80] provides an adequate framework for their future standardization. Local network vendors may then provide Network Layer or Transport Layer protocols which follow the standards.

Voice is analog in nature, fitting into a fixed frequency range. A fairly wide range of network voice processing may exist, from the extreme of intercom systems that share transmission media with computer networks, to complex network processing which synchronizes voice, video, and data into teleconferences while receiving and transmitting to distributed points.

Digitized voice today is currently offered primarily by PBX vendors, though it seems likely that future non-PBX local networks will also provide a voice capability. Since voice has special delay (and looser reliability) requirements, different network and transport-layer protocols are required to adequately support voice applications.

Video information can be digitized and transmitted through a local network. Alternatively, if the local network provides analog channels as well as digital channels (e.g. a broadband cable system), then the unprocessed video signal can be transmitted. Digitized video requires high data rates.

Digitized video generally does not require bits to be transmitted reliably. If the video image is for "real-time" use, the delays required for bit reliability may not be tolerable. Consequently, video transmission not only requires high data rates, but also requires connection-oriented protocols with expedited services and differing quality of service parameters (flow and error control) than connection-oriented protocols intended for data transmission [BURJ80] [BURJ81a].

The necessity for video services has has a large impact on the resulting network procurement.

Imagery includes facsimile, raster graphics, and slow-scan video with large blocks of data generally transferred. Delay requirements are minimal compared to voice or real-time video, and reliability requirements are not as strict as those for file transfers or terminal traffic. Image traffic requires protocol features within the transport [BURJ80] [BURJ81a] and network layers.

The type of information transmitted across the network - data, voice, video, or images - impacts the required network services in two major ways:

The network must be capable of handling the data rates typical of the traffic type to be supported.

The network must provide the Network, Transport, and Presentation Layer protocols appropriate for traffic type.

<u>3.1.7 Basic Network Services.</u> A common set of shared services are provided to the network node by the communication protocols specified according to the ISO Reference Model [ISOR80], as described in Section 2.3.1 Services of primary concern are:

1. Transmission of single transactions, for network control and user applications.

2. Initiation, termination, and control of switched point-to-point virtual circuits.

3. End-to-end flow control between network nodes on a virtual circuit, to ensure that data is not lost during momentary transmission facility overloads. This includes speed matching of device data rates between source and destination nodes.

4. Error control to ensure accurate information transmission, with varying bit error rates achievable.

5. Code translation between nodes, based upon interface requirements.

6. Terminal translation between otherwise incompatible nodes, such as between word processing and graphics terminals.

A large number of applications require the reliable transmission of small, independent pieces of data called "transactions". Typical of these applications are database access, network monitoring, process control, and financial data entry. These applications do not require the transfer of a large quantity of data in sequence, and in the case of real-time and interactive applications, cannot tolerate the delays required to establish a connection.

These applications require the features of transaction protocols that manage the transfer of independent data blocks. Transaction protocols are connectionless protocols residing at the transport layer of the ISO Reference Model, and do not involve explicit "connection establishment" and "connection termination" phases.

If the installation's applications require transactions, the local network must either support such a service, or additional user software will be required within hosts.

Virtual circuits are the most common mechanism for transferring data across a network. A virtual circuit establishes a connection between two session, transport, and network entities [BURJ80] [BURJ81a] [BURJ81b] which allows them to exchange data reliably and in-sequence across a network. The circuit must be explicitly established. During circuit establishment the appropriate tables and state information are set up in the participating nodes; this allows proper sequencing and acknowledgments to occur.

Because each network interface unit may handle data at variable receive and transmit rates, congestion may slow down or stop the flow of information. Therefore, it is necessary that the network and its hosts have some method of flow control that maintains the movement of traffic.

If the capacity of the lines and nodes is always sufficient to carry the load, the free flow of traffic might still not be guaranteed. Exit of packets from the network at a destination might be impeded because the network cannot deliver packets faster than they are accepted by the destination node.

The network's carrying capacity must be planned to cope with expected demands, but events may cause the capacity to be exceeded. Economics may prevent the provision of transmission capacity for any conceivable load variation. Statistical variations of traffic, surges of traffic due to outside events, or planning which did not properly anticipate the extent of demand will also cause occasional peaks in excess of network capacity. When devices of differing speeds communicate, the network must provide buffering and flow-control services that implement "speed matching" between source and destination nodes.

Two primary types of flow control exist: packet network flow control and congestion avoidance. If part of the network becomes overfilled with packets it becomes impossible for packets to move. This is called "congestion". "Flow control" regulates flow in normal operation and is principally a method of transmitting flow restrictions back to the place where the flows can be controlled. Congestion avoidance prevents overloading of the network.

Flow control may be provided throughout multiple layers of the ISO Reference Model [ISOR80], as appropriate to the functions each layer performs and the services provided to the next higher-layer protocol entity. Each layer allows selection of quality of service parameters related to throughput, transit delay, and connection set-up delay, and performs a flow control function to regulate the flow of data between two protocol entities on a connection.

For example, the transport layer performs end-to-end flow control on individual transport-connections. In addition, the expedited data transport service may be selected by a session-layer entity to bypass the flow control of the normal transport data unit, where the requisite transport function uses a flow control mechanism different from that for the normal case.

Error control services provided by the network may include guarantees on delivery of packets, guarantees on the bits within the packets, or guarantees on the sequentiality of the received data. The degree of error control provided is a function of the protocol being used. A connection-oriented transport protocol provides sequenced delivery, bit, and sequence error control [BURJ80], whereas a network-layer datagram protocol may not guarantee bit reliability.

Error control is provided throughout all layers of the ISO Reference Model [ISOR80], as appropriate to the next higher-layer protocol entity to which services are provided and the functions the layer performs. Each layer provides selection of quality of service parameters related to error control, and error notification to the protocol entity of the next higher layer when it detects any unrecoverable error.

As an example, the transport layer provides end-to-end error control to the session layer. As an establishment service, the transport layer allows selection of guaranteed values of residual error parameters:

1. Detected but unrecovered errors.

2. Undetected errors, arising from alteration, loss, duplication, disordering, and misdelivery of transported data.

Functions performed by the transport layer include monitoring of the quality of service, end-to-end error detection, and error recovery from detected and signalled errors. Thus any requirements on the network's error control service generally will translate into requirements on specific network-provided protocols.

Often, it is necessary for devices using different coding schemes (e.g. ASCII and EBCDIC) to communicate. This can be handled by translating from one code to Network Normal Form at the source, then translating from that form to the second code at the destination. Such a service is placed in the Presentation Layer of the ISO Reference Model.

A more difficult translation problem than simple code conversions involves terminals of vastly different structures - for example, communicating word processors and graphics terminals. Translating between such disparate devices requires sophisticated software, as handled by a Virtual Terminal Protocol [SHEC81] and Data Presentation Protocol [CLOS80a] [CLOS80b].

A terminal translation service is not anticipated to be provided by many local vendors, and if it is a strong network requirement it will most likely require custom software.

3.1.8 Higher-Layer Functions. Most vendor networks provide a subset of the Basic Network Services described in the previous section. Additional services for specialized application may be required. These include word processing, file transfer, facsimile translation/distribution, electronic mail, teleconferencing, electronic funds transfer, and file archiving.

If the installation requires that the procured network provide one or more of these higher-layer services, the requirements analysis becomes more involved. Different vendors have different solutions for providing the functions mentioned above; unless a detailed set of requirements is determined, comparisons may be meaningless.

3.1.9 Security and Privacy. Network privacy becomes a requirement when multiple user groups access the network and the information maintained by each group must be kept from others. In addition, the network should be kept safe from unauthorized attempts to access the network. The objective of a data network's security mechanism is to protect users from misdelivery of data. Misdelivery could result from network error or a deliberate attack. Security considerations include:

- Security.
- Access authorization and protection.
- Encryption.

3.1.10 Remote Communications. Users and devices within the network may require communications to outside networks. Internetworking includes communications to other local area networks or to public data networks.

Communications to local area networks with common administrative and operational services is straight forward; often protocol translation is not required. Simple gateway devices or "bridges" are sufficient. Communication with different local networks requires protocol translation obtained from sophisticated gateway devices. Local network connections to public data networks use gateway devices that implement local network protocols on one side and X.25 protocols on the other. X.25 is an interface specification between terminals operating in the packet mode (in this case the local network) and public the data network.

3.1.11 Establishing Service Requirements. The installation administrator performs the following analysis to determine the services required of the local network:

1. Examine the FIPS register and other sources for applicable current and planned standards so they can be incorporated by reference.

2. Catalog the intended applications. Include, for example, real-time graphics, terminal/host access for program development, word processing and database access.

3. Catalog devices that have direct network access. Include the method of attaching the device and the device's interface.

4. Catalog the services that network applications expect. For example, a file transfer application requires a connection-oriented service from the local network's transport protocol; database access may require connectionless support.

5. Catalog applications that will not be implemented in-house.

The service requirements analysis states which devices, interfaces, basic protocol services and higher layer functions are explicitly required.

3.2 Network Traffic Characteristics

The volume and arrival rate characteristics of the traffic between communicating equipment within the network determine many aspects of the required network's design. Cataloging these characteristics involves an analysis of the node-to-node traffic flow, and includes collection of information concerning the following areas:

1. Traffic throughput requirements

2. Traffic class - streams and transactions

3. Tolerable network delay

4. Interconnectivity between nodes

5. Concurrency of messages and circuits

The following sections address these areas in detail.

3.2.1 Traffic Throughput Requirements. Determination of traffic throughput requires knowledge of transmission medium throughput and node throughput characteristics. Determination of these characteristics involves the collection of application traffic statistics, in areas such as:

Peak, average, and maximum data rates of devices and links in the network.

Sizes of data blocks (messages) transferred between network nodes.

Each user device on the network may support a number of separate applications requiring node to node communication. These applications generate and consume traffic according to their specific needs. Traffic patterns into and out of each user device are estimated to determine the required speeds and capacities of the network and its nodes.

An application's traffic characteristics can be analyzed using two variables: message size, and interval between messages.

Both variables can be statistically described in terms of peak values, average values, and variances. For the purposes of network capacity planning, simple characterizations are sufficient. More formal methods may be used as the agency deems necessary, such as the Poisson distribution techniques described by Tobagi [TOBF78].

For example, connectionless database access from a terminal node may generate messages of average size 20 bytes, with peak generation rate of once every 15 seconds. In contrast, an application requiring connection-oriented file transfers may generate 512 byte blocks at a steady 100 blocks per second.

The traffic exchanged between a user device and the network is the sum of the traffic of that device's individual applications. Although the total traffic involving a single device can be extremely complex, it is generally sufficient to describe its average and peak data rates (transmitted and received). Depending on other requirements such as delay or reliability, the network's nodes and links can be designed to handle peak or average traffic.

<u>3.2.2 Traffic Class - Streams and Transactions.</u> Besides the raw data rates and message sizes exchanged between applications, the "class" of traffic has an effect on the network's design. Data traffic falls into two basic classes:

Stream (connection-oriented) traffic, in which large quantities of data are transferred between two nodes.

Transaction (connectionless) traffic, in which small amounts of data are transferred in independently handled blocks.

File transfers, word processing applications, and voice are typical of stream traffic. These transfers are managed by the services of a real or virtual connection; these connections require a connection establishment phase prior to the transfer of any data. During connection establishment, network resources such as buffers are initialized.

Database query/response and process control applications are typical of transaction traffic. These applications do not require the transfer of large amounts of data, and generally cannot tolerate the overhead involved in opening a connection. Separate transactions may arrive at the destination in a different sequence from that in which they were sent.

The difference between stream and transaction traffic does not affect transmission capacity planning as much as it affects the processing requirements of the network nodes. The most suitable protocols for a given installation depend on the estimated stream/transaction traffic mix. The ICST transport protocol [BURJ80] provides the services required by both classes of traffic. For example, if a vendor's network is intended to only support transaction applications, a user's stream applications will be handled poorly unless additional protocols are implemented within the user's own devices.

3.2.3 Tolerable Network Delay. Another type of traffic requirement that differs among applications is the amount of network-induced delay which may be tolerated by the application. Requirements on delay take the form of a minimum value for the delay experienced by data during transport across the network. Requirements on delay variance take the form of a minimum value on the variance experienced by individual packets within a data stream.

Applications can be classified into three categories with respect to their delay requirements:

1. Those applications that can not perform their function if network-delay exceeds a certain minimum value. Typical of this category are real-time process control applications.

2. Those applications that can tolerate some delay, but which function best if the delay is kept under a certain minimum value. Typical of this category are "interactive" applications involving user terminals, where user satisfaction drops as delays increase.

3. Those applications that are insensitive to network delay. This includes batch processing, file transfers, and electronic mail.

Some applications have communication requirements that do not fit neatly into one of these categories. For example, an application that occasionally requires "priority" transmission of crucial control information. Applications in the third category function best if delays are minimized.

The total delay is composed of many separate network and non-network elements. For example, delays in a database query result from:

1. queuing delay at the query's entrance into the network.

2. processing delays as protocol interpreters manage the query's transmission.

3. transmission delays through the network's media.

4. queuing and processing. delays internal to the network within store-and-forward nodes

5. processing delays within the database server as the response is generated.

6. similar queuing, processing and transmission delays as the response traverses the network back to the user.

Not all of these delays are network-induced, and it is often difficult to determine which component is responsible for excessive delays.

The above discussion is applicable to any networking environment, long-haul or local. Restricting discussion to local networks, the following is noted:

* In a local network, the major component of network-induced delay comes from protocol processing rather than from transmission delays through the media [CLAD78] [TOBF79].

This generalization is often viewed as the most important distinction between local and long-haul networks, since in a long-haul network the transmission delays tend to dominate the total delay.

The impact of this generalization on local network design and procurement stems from its emphasis on computational aspects of the network over the communication aspects. For example, a local network with data links running at 2 Mb/s will not automatically have twice the throughput of a similar network with 1 Mb/s links, since the throughput bottleneck in both cases will most likely be protocol processing rather than transmission through the links.

Some applications have requirements for minimal variation in the delay experienced by the separate packets within a data stream. Packetized voice is a good example. If individual packets experience widely varying delays through the network, the reconstructed signal at the receiver will be unintelligible.

Delay variations can be eliminated through one of two techniques:

Received segments can be buffered at the receiver into a queue, then "played out" to the user at a constant rate. This minimizes delay variation at the cost of increased delay.

The data stream can be allocated enough dedicated transmission resources that individual segments automatically receive exactly the same network service. An example is time division multiplexed links. This minimizes delay variation at the cost of decreased total throughput.

3.2.4 Interconnectivity Between Nodes. A single node may house multiple applications that require network access to a variety of other nodes. The connectivity requirements among the different nodes is included as part of the general catalog of network traffic characteristics.

Topology independent data flow diagrams can be developed as described in Section 4.1. These graphically define the required traffic between nodes. The diagram consists of points representing the nodes with arrows between nodes labelled with data rate, delay and stream or transaction traffic requirements. Separate data flow diagrams for average and peak traffic should be included. With such a diagram, an analyst quickly determines the number of transactions per second and the number of simultaneous stream connections per second that a particular node must manage. These flow diagrams are also useful in selecting an appropriate topology for the network. Determination of the network topology must also take other issues (Section 2.2) into consideration.

3.2.5 Concurrency of Connections. Each node must sustain the sum of the traffic processing requirements of its applications. The communications hardware and software used by a node must have the necessary capacity.

Network resources are allocated to provide the required capacity. Virtual and physical circuit resources require interfaces, buffers and state tables. These resources must suffice to meet the needs of all circuits that exist during periods of maximum connection concurrency. The maximum and average number of concurrent connections at each node should be specified.

3.2.6 Establishing Traffic Requirements. The procurement of a local network includes a set of capacity requirements based on anticipated traffic. The following traffic characteristics are determined through analysis of the intended applications:

1. peak and average rates, for stream and transaction traffic.

2. Peak and average totals for stream and transaction traffic per connection.

3. delay and delay variance requirements.

4. concurrency of connections for each node.

5. aggregate network throughput.

3.3 Reliability Requirements

The required reliability of a network and its components varies depending on the network application. Requirements are determined analyzing following areas:

1. Reliability of individual network services.

2. Availability.

3. Node or link failure.

4. Bit errors, packet errors, retransmissions.

5. Placement of network control equipment (centralized or distributed).

The following subsections discuss these areas in detail.

3.3.1 Level of Network Services. A series of examples illustrate reliability requirements for file transfer, database query, voice connections and character terminals.

File transfer exhibits three reliability requirements: Delivery -- everything sent must be received; Sequencing -- everything received must be in the same order in which it was sent; and, Correctness -- everything received must have no bit errors

Database queries exhibit different requirements. Suppose users at a number of interactive terminals simultaneously query a remote database. Each user requires that his query be answered; but, the order in which they are answered is immaterial. Responses from the database may not even be answered in the same sequence as the original queries. In such an application, high bit and delivery reliability are required, but sequence reliability is not necessary.

Voice applications tolerate bit errors since the ear smooths over a scratchy received signal. Human speech is already so redundant that it makes little sense to demand absolute bit perfection within a digital voice stream. A cavalier attitude towards bit errors in a voice stream may cause problems if a highly encoded voice digitizing technique is used since a single bit error may invalidate an entire digitized phrase.

Typical applications using asynchronous character-at-a-time terminals involve echoing from the remote host. In such a configuration, errors in transmission will be noticed by a human operator. In general, the same reliabilities required for file transfers are also desired here. However, the presence of a human removes the requirement for high levels of reliability provided by the network.

3.3.2 Availability. Different applications have different reliability requirements for available network services. A node or link failure most affects the availability of these services.

Requirements for reliability are determined by identifying critical elements of the system and by defining Maximum allowable time and frequency of outages that can be tolerated.

The system's critical elements range from those nodes that provide a network service critical to the proper functioning of other nodes such as a name server, to those nodes that are ordinary from the network's viewpoint but which provide the network access for some critical application. Network links can be classified according to their importance to the network or application. Failure duration and frequency can also be determined on a per-application basis.

Many techniques are available to enhance the reliability of a network in the event of a node or link failure. The most basic approach is to design the network with as few critical nodes and links as possible. For example, a network with a central controller responsible for allocating transmission resources (e.g. a TDM channel allocator) is completely dependent upon the proper functioning of that node. To eliminate the reliability problems inherent in such systems, a variety of "distributed control" mechanisms have been developed that allocate link transmission capacity without the single failure point of the centralized controller. Examples of such distributed control techniques are the variations on "carrier sense multiple access/collision detection" (CSMA/CD) used in many contemporary coaxial cable networks. Each node independently determines when it should transmit on the shared transmission medium.

The second technique for managing a network in the face of node or link failures involves redundancy of the critical elements, allowing operations to continue by switchover to backup elements. With sophisticated network control protocols, switchovers are made transparent to user applications. For example, adaptive routing techniques in mesh topology networks incorporate automatic switchover to backup links and nodes directly into the normal routing mechanisms of the network.

<u>3.3.3 Establishing Reliability Requirements.</u> Reliability requirements are application-specific, and the network-wide reliability requirement will be the requirement of its most important application. Consequently, a determination of the installation's reliability requirements should begin from a list of intended applications. Each one has requirements on delivery reliability, bit reliability, and sequence reliability. The impact of these requirements are in the protocols offered by the procured network.

3.4 Network Growth

The data processing environment is dynamic, and the network that provides the capability for data communications within that environment must be capable of meeting the changing needs of the target installation. Part of the dynamics involves network growth in the following subareas:

- Addition of nodes to existing networks.

- Additional types of traffic.

- Changes to existing types of traffic.

- Internetworking modifications.

- Movement of equipment.

The following subsections discuss these areas in more detail.

3.4.1 Addition of New Network Nodes. Anticipation of the addition of new nodes to the network over a specified period of time is an important consideration during the requirements definition phase. There are two distinct reasons why new nodes are added to an existing network:

1. additional processing capacity for applications is desired. This increases the total network traffic.

2. redistribution of centralized applications into distributed processing environment. This may enhance reliability and software modularity, and may decrease total network traffic.

As an example of the second type of network expansion, consider the case of interactive terminals accessing a timeshared computer on the network. By adding new nodes in which users can do local editing, total network traffic decreases.

Although it is difficult for an organization to predict its requirements for addition of new nodes, considerations can be included in the analysis as early as possible. Aggregate network traffic capacity, network node control and the ease of introduction introducing new nodes are features of a local network that impact the ability to grow.

3.4.2 Addition of New Traffic Types. Addition of new types of traffic after the network has been operational for a period of time can be considered while the requirements are initially defined. New types of traffic may require significantly different transmission methods and/or protocols. For example, digitized voice is a different type of traffic from bursty data streams, which in turn is different from transaction traffic. The addition of new types of traffic to an existing network may require enhancements to the network's physical and protocol architecture.

For example, it is not feasible to use standard connection-oriented service to support voice connections. Voice traffic has particular delay and reliability requirements that differ significantly from those of data streams. A common protocol has been developed by ICST [BURJ80] [BURJ81a], but this is typically not yet available in vendor products. Instead, separate sets of protocols are used, or even separate physical links, for the different traffic types.

3.4.3 Changes to Existing Types of Traffic. As time progresses during the life of the network, traffic characteristics may change, reflecting modifications in installation operations. Traffic volume and traffic generation statistics may both evolve. This requires designing adequate capacity into protocol processing nodes and transmission links.

In addition, the protocol interpreters should offer a flexible service that can be adjusted to the new traffic characteristics. As the traffic characteristics change (due to changes in applications), parameters such as timeout values may no longer provide the most efficient transmission. Such protocol parameters should be settable. This allows the network administrator to fine tune the network allowing for changing traffic conditions.

3.4.4 Internetworking Modifications. Modifications of internetworking capabilities, such as communicationing to additional networks or increasing traffic to existing networks, occur as new networks are installed. It has been the experience of many local network installations that although internetting requirements may not be initially present, sooner or later it is desirable to allow internetwork communication.

3.4.5 Movement of Equipment. Considering the dynamic environment in which many networks exist, it is often necessary to move equipment between offices or buildings. Flexibility in relocating equipment is improved when the network has the following two characteristics:

1. Physical attachment to the network is easy and inexpensive.

2. Network control mechanisms do not depend on the physical location of particular nodes.

For example, some local networks allocate transmission grants based on a node's physical location (similar to daisy-chain peripheral busses). With such an architecture, the performance of the network as seen by a node will vary with its physical placement. This may hamper the flexibility with which equipment may be located.

Node addresses (used to route data to its designated destination) may depend on a node's physical location in some network designs. This requires the changing of addresses whenever a node is moved - and all other nodes must similarly be advised of this change of address.

During the network requirements definition phase of a procurement, the likelihood of equipment relocations must be considered. Otherwise the network's architecture may preclude a flexible equipment location capability.

3.4.6 Establishing Growth Requirements. All local network installations change as their applications mature; these changes can be anticipated in the initial procurement.

The following growth requirements should be considered:

1. Growth through the addition of new network nodes increases traffic, requiring an adequate capacity in the links and processing nodes.

2. Growth through the addition of new traffic types require new protocols.

3. Growth through interconnections with other networks require an internetting protocol architecture from the very beginning.

Once the most likely scenario is established, growth requirements can be translated into requirements for network capacity and protocol architecture.

3.5 Maintenance Requirements

A local network's maintenance costs may dominate the total life-cycle costs of the installation. Each installation requires its own approach to network maintenance. Techniques used to maintain the network must:

- ensure a level of continuous operation commensurate with the fundamental needs of the network's applications,

- provide service within the installation's financial and personnel resource limitations

These crude requirements on the network's maintainability can be refined into specific network requirements. Three network requirements that determine the basic approach to maintenance are:

1. Maintainability of network software and hardware components

2. Compatibility with established or de facto standards

3. Network performance monitoring capability.

3.5.1 Hardware and Software. The individual hardware and software components of the network must be maintained. For hardware this generally means preventive and corrective maintenance to ensure a working system. The term "software maintenance" implies changes that enhance or modify a network function.

When specialized network capabilities are required, it is necessary to development non-standard software by either the network vendor or personnel internal to the target installation. Inclusion of such specialized software affects maintenance.

The ease with which custom software is maintained depends on a variety of conditions, including:

1. modularity of the design

2. use of accepted programing practices such as structured programing techniques

3. quality documentation

Benefits anticipated from the use of custom code should be carefully weighed against the potential problems of maintaining the code. Network vendor code has the same maintenance problems if not developed according to good software engineering standards. The network procurer has to be wary of buying sophisticated software whose full capabilities will not be useful within the target installation. Even if the basic price of such a system seems reasonable, maintenance problems may be larger, and total life-cycle costs may balloon.

Depending on the technical sophistication of the installation's staff, software maintenance can be performed in-house or contracted out to the network vendor or a third party.

The potential advantages of in-house software maintenance include:

1. quick response to problems

2. administrative control over individual maintenance decisions

3. enhanced technical control over total network functions

4. reduced total maintenance costs

In-house software maintenance is cost-effective when the staff is technically competent and the software is carefully documented. If these criteria are not met, in-house software maintenance leads to exploding costs and system deterioration; often, simple software modifications affect other functions in unexpected ways. On the other hand, if the software was originally developed in-house, then in-house maintenance may be cost effective.

During network procurement, the costs of outside software maintenance and the costs of unscheduled network outages must be weighed against the risks of in-house software maintenance to determine the installation's maintenance requirements.

Maintenance of the network hardware is considered from two perspectives: periodic preventive maintenance, and correction and recovery from failures. The division of maintenance responsibility is structured along these lines, with periodic preventive maintenance done in-house, and with outside maintenance contracted if the network fails.

One problem of a complex hardware/software system such as a local network is the determination of which component or combination of components is responsible for network failure. Often, it is difficult to determine whether the network failure resulted from a hardware or software malfunction. Thus the maintenance of network hardware cannot be performed in isolation from the maintenance of network software.

3.5.2 Compatibility with Established Standards. Less expensive off-the-shelf solutions is the primary benefit gained from procurments of equipment that comply to standards. Compatibility also foster competition allowing proposals to include multi-vendor sources of equipment.

As the network market matures, and as national and international standards are adopted, the local network procurer will structure his system in a way that promotes vendor independence.

The advantages of a vendor-independent system from the viewpoint of maintenance include:

1. the ability to initially choose equipment for maintenance reasons (e.g. good vendor service reputation) rather than being "forced" to use a particular vendor.

2. the existence of "second-sources" provides insurance against failure of the primary vendor.

3. ability to drop a vendor if a record of poor maintenance is established.

3.5.3 Network Monitoring. The use of monitoring equipment may be helpful in controlling maintenance costs. The specific role of a network monitor varies depending upon the network's underlying technology. The following functions are common to many monitoring systems:

1. centralized statistics gathering on error rates and utilization of each link or node.

2. automatic failure detection on critical system elements.

3. artificial traffic generators to probe element status or exercise specific network functions.

4. frequency-response monitoring of analog network components

Equipment to perform the above functions can be provided by the network vendor or maintenance contractor. Installation personnel can then be instructed on its proper use, allowing problem causes to be identified prior to the arrival of maintenance personnel. Early problem identification can allow quick fixes such as board replacement. Often, the initial cost of this equipment is recovered through reduced total maintenance costs.

<u>3.5.4 Establishing Maintenance Requirements.</u> The network procurer identifies the installation's requirements with respect to maintaining the network once it is operational. A primary consideration is the cost of maintenance versus the cost of initial system procurement.

The following questions should be answered in determining the network's maintenance requirements:

1. Do the anticipated network applications require custom hardware or software? Custom interfaces? Or do budget and in-house resource limitations require the use of standard "off-the-shelf" network products?

2. What requirements do the anticipated network applications have with respect to mean time to failure and mean time to repair?

3. Can the installation's staff be responsible for maintaining parts of the network, or must this be done by vendors or contractors?

4. Do international or de facto standards meet the applications' requirements for functionality and performance? If so, can a vendor-independent system be procured and maintained?

5. What network monitoring equipment should be incorporated into the network? Such equipment must ultimately pay for their initial costs through reduced maintenance costs or through enhanced network performance and utilization.

3.6 Requirements Summary

The procurement of a local network begins with a detailed requirements analysis. The analysis is driven by the applications to be supported and the analysis determines which proposed networks are solutions and which may fail to adequately support the installation's needs.

Services expected of the network are the primary requirement. The protocols of the system are determined by the services the system supports. Traffic characteristics of the anticipated applications are analyzed to yield estimates of the transmission and processing capacity of the procured network. Reliability, growth, and maintenance are also important requirements.

4. SOLICITATION PREPARATION PROCESS

Given the features provided by contemporary local networks and the analysis of specific requirements for the target computer installation, network procurement requires detailed preparation of a network solicitation document per Government procurement regulations [FEDE80].

The goal of this section is to provide guidance in preparation of network requirements within the work statement of the solicitation, as applicable to the technology of local area computer networks. The section is not intended to present standards for format and content, but rather to describe the specification process for network requirements in terms of function, performance, and environment.

The solicitation requirements should be written in clear, concise language that is easily understood by vendors. The importance of well-written solicitation documents cannot be over-emphasized, since they express the requirements in meeting agency goals. Misunderstanding can be a significant factor in contract negotiation and vendor performance; therefore all aspects of the services and/or items to be delivered should be understood by both agency and prospective vendor. In general, the solicitation should contain the following:

Scope and Objectives:

Every solicitation contains introductory paragraphs that present a clear description and understanding of the overall scope and objectives of the procurement. Scope identifies the major elements of required work and the end result or product desired. The manner in which the Scope is defined also governs the amount of direction that the Government gives and that the vendor accepts during the life of the contract.

General Background:

General background information includes a brief history, relationship of the effort to other procurements, technology to be used or not used and a list of reference documents. This information provides vendors with the general objectives of the federal agency that keep delivered work and/or products in proper perspective.

Vendor Tasks and Deliverables:

The solicitation provides detailed descriptions of the analysis and studies to be performed, the services to be provided, the equipment and software to be delivered, and the contract management and review systems employed. The requirements are described in complete detail, whether by direct statements or reference to full or partial specifications and standards. Proper language is used to make multiple interpretations impossible. The specific purpose must be kept in mind, stating the results required and not methods the vendor must use in performing necessary work.

In making judgements concerning whether material is included in a solicitation, the following questions may be asked:

1. Is it necessary in order to accomplish the effort and/or provide the best products?

2. Does it tell the vendor what he is required to provide?

3. Is it necessary in order for the vendor to determine what is required for contract completion?

4. Is there a method to determine when the basic task is complete, i.e. can it be priced?

Material or tasks that do not pass these tests are generally redefined or left out of the solicitation.

The specification of requirements and vendor tasks for the local area computer network requires three areas of concentration: system description, statement of network requirements, and general network requirements specification. When preparing the solicitation the agency remains aware that the intention is for prospective vendors to respond with proposals and/or quotes containing the appropriate features highlighted and for obtaining a sufficiently wide range of competitive responses, allowing choice of the best approach.

It is important that the procuring agency describe in detail the system within which the network operates. The local network itself does not offer the total system solution to the application, but provides communication capability to system entities. The network may have to fit within an existing operational system or may be procured in parallel with other system subsets to provide a new environment. In either case, a clear delineation of system and network responsibilities should be made to ensure that the bids received are responsive to the problem.

Given the environment within which the network operates, a detailed statement of the individual network functional and performance requirements may begin. The areas to be specified are:

- o The network services that must be performed for users and system equipment.

- The characteristics of the traffic between communicating entities.

- The reliability expected of the network as a whole and its components.

- The growth in function and performance over the specified time period.

- The maintainability of the network to ensure a level of continuous operation commensurate with the fundamental needs of the network's applications and within the resource limitations of the target installation.

The local network to be procured may be considered as a subsystem within the total system previously described. As such it is a functioning entity with the following system-level requirements placed upon it:

1. Life-cycle of operation, such as phased introduction of capabilities, installation, acceptance, operation period, and maintenance.

2. The particular criteria used for accepting the network as a whole, with test and verification requirements.

3. Documentation that must be delivered, including financial, administrative and management, and technical data throughout all phases of the contract.

4. Procedures for modifications to the network after review and acceptance points.

4.1 System Overview

The solicitation document provides network vendors with an understanding of the overall system and environment within which the network operates. The first section presents a system overview providing vendors with a more complete understanding of what is required. This helps them better explain how their products can best satisfy the networking requirements. The system overview also helps the federal agency verify that it has a system-level understanding of the network's functional responsibilities. The system overview consists of the following:

1. A description of what the system does and where the system fits into the overall agency organization and its function.

2. A functional description of the subsystems that make up the system and how the network is to support these subsystems.

3. A configuration diagram of the major components of the system and their functional responsibilities.

4. A data flow diagram showing the major components of the system and other connected systems.

5. A classification and characterization of the users of the system.

6. Operational characteristics of the system including reliability, performance, loading, and growth.

4.1.1 System Application. The primary application or purpose of the system is briefly described. The purpose and organization can be described as in the example that follows:

The primary application of System X is to provide management planning information to Group R of Organization Y.

Group R is the element of Organization Y responsible for four year cycle manpower and resources planning.

Organization Y has full forecasting responsibilities for Program Z of the XYZ Agency.

A hierarchical representation is often appropriate to indicate where the system resides in the organization as a whole. Whenever possible these structure diagrams show functional as well as organizational structure.

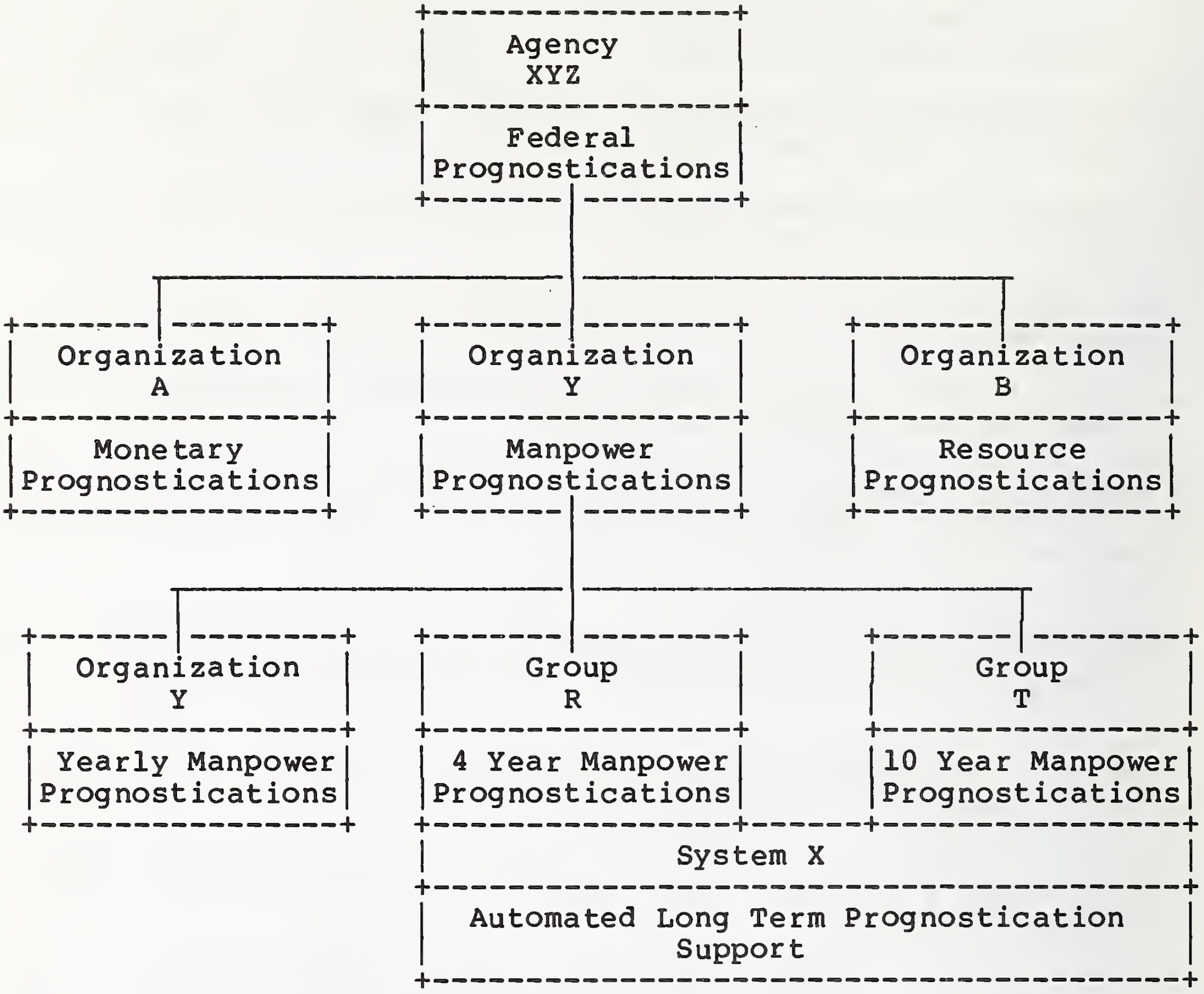

Agency XYZ
Federal Prognostications
Organization A
Monetary Prognostications
Organization Y
Manpower Prognostications
Organization B
Resource Prognostications
Organization Y
Yearly Manpower Prognostications
Group R
4 Year Manpower Prognostications
Group T
10 Year Manpower Prognostications
System X
Automated Long Term Prognostication Support

Figure 4-1. Hierarchical Structure Diagram Example

4.1.2 System Services. System services are capabilities provided by the system to its users in support of the system application. These services are briefly described.

EXAMPLE:

System X provides the following services:

1. Interactive support to forecast analysts' workstations of Group R.

2. Maintenance and control of the manpower and resources database of Organization Y.

3. Off-line storage and production services for Group R and Group T.

The relative priority of each of the services are discussed in order to give the vendor an understanding of operational priorities. The functional subsystems can also be used to group services.

EXAMPLE:

The functional subsystems of System X are:

- Analyst Support Subsystem,
- General Support Subsystem,
- Database Subsystem,
- Production Subsystem, and
- Storage Subsystem.

4.1.3 Equipment Configuration. The system equipment configuration is presented as a general block diagram depicting the physical layout of the building(s) housing the system. General equipment types are identified. Only the identification of functional subsystems to physical components are illustrated.

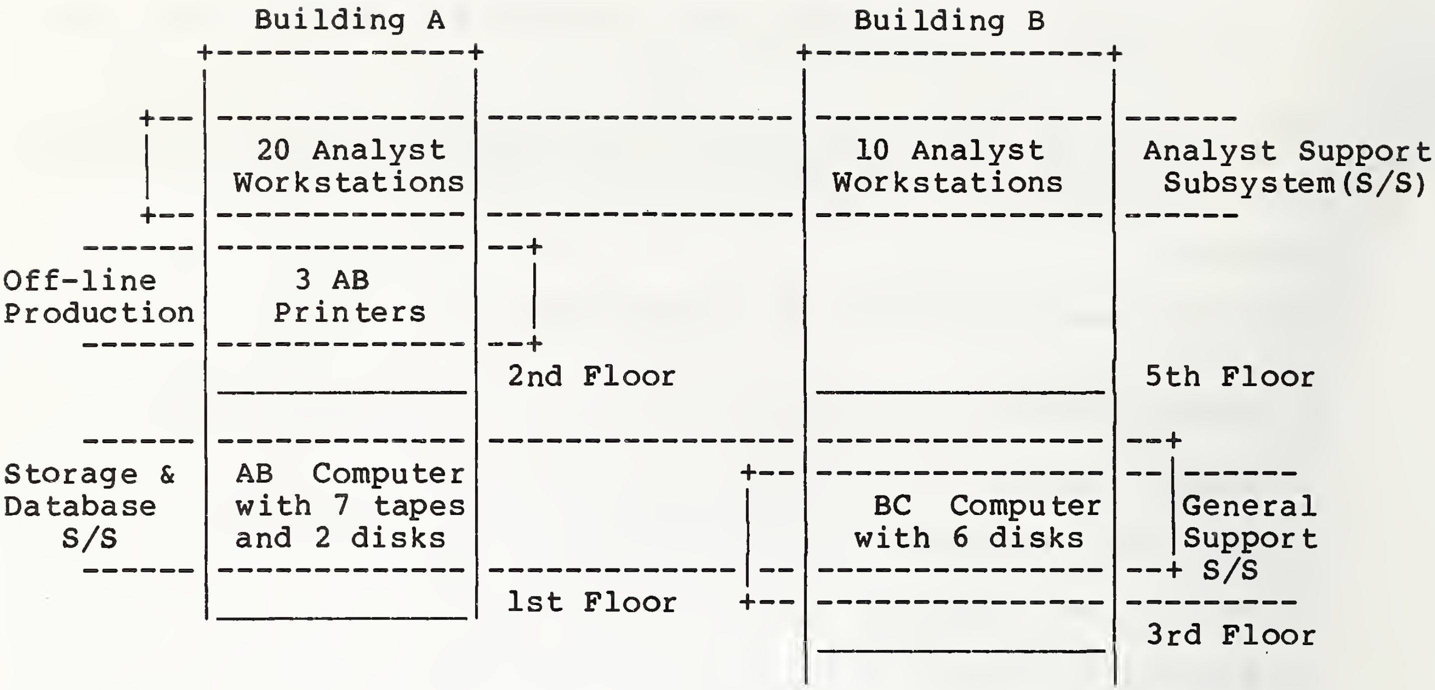


Figure 4-2. System Functional Configuration Diagram Example

Physical layout diagrams of all offices, equipment/computer rooms, floors, and buildings that house pieces of the system are provided. This information gives the vendors a physical perspective of the system and essential information required for network layout. Equipment in the diagrams are associated with functional subsystems.

4.1.4 System Data Flow. Illustrations of the system data flows, including those not supported by the network, provide an understanding of how information moves through the system. The system data flow diagram depicts the flow of information among the functional subsystems and external sources at the system level, but is not as specific as the network data flows described in Section 4.0, Solicitation Preparation Process.

4.1.5 System Operations. System operations describe the sequence of activities that are supported in a normal operational period, defined as a recurring sequence of system activities (e.g. 6-hour data base updates, daily analyst reports, weekly group forecast). System operations should be described to vendors for understanding of how the system is used and for help in identifying when and why peak traffic loads occur. A representation of when and the approximate amount of data transferred in a particular operational sequence can be illustrated with a Data/Sequence chart.

```
+-------------------------+   +---------------+    +---------------------+
| Manpower and Resources  |   |General Support|    |Reformatted Update   |
|Updates From Organization|-->|S/S Update Data|--->|Data Transfer to     |
|A  (5 2500 byte records) |   | Reformatting   |   |   Database S/S      |
+-------------------------+   +---------------+    |(5 2200 byte records)|
                                                   +---------------------+
              Manpower and Resources Data Update Sequence
                      Repeated Every Six Hours

Time        0500                       0505                      0510
```

Figure 4-3. Data/Sequence Diagram Example

4.1.6 System Users. The users of the system are categorized in terms of how they use the system and by the characteristic functions they perform. These user characterizations help vendors determine the type and level of interfaces required. For example: "A forecast analyst shall use a personal workstation 5 hours a day. The analyst shall make connections to the various functional subsystems with which communication is required. The analyst is well versed in forecast algorithm development but not the particulars of system/computer operation. A user friendly interface that gives the analyst status and helps resolve problems or questions is required."

4.1.7 System Characteristics. Various system characteristics provide useful insights into the operational system environment. These characteristics include system performance, reliability, availability, growth, and loading.

System Performance

The specification of general system-level performance requirements between functional subsystems can be expressed in terms of the maximum (Tmax) and minimum (Tmin) time required for a particular information exchange between two subsystems. The computation of these values requires estimates of the maximum and minimum processing time to fetch the data from the subsystem where it is stored (e.g. memory, disk, or tape) and transfer it to the network (tmaxA and tminA), to transfer the data across the network (tmaxNET and tminNET), and for the receiving subsystem to accept and store the data (tmaxB and tminB). These values can then be summed to give estimates of the maximum or minimum performance time.

$$\mathrm{Tmax} = \mathrm{tmaxA} + \mathrm{tmaxNET} + \mathrm{tmaxB}$$

$$\mathrm{Tmin} = \mathrm{tminA} + \mathrm{tminNET} + \mathrm{tminB}$$

The values used in this equation for tmaxNET and tminNET are approximations, less accurate than the values obtained in the detailed network analysis in Section 4.2.

Reliability/Availability

Reliability is defined as the probability that a device will perform without failure for a specified period of time or amount of usage. System reliability is based on the composite reliabilities of all the components of the system. System availability is expressed as that portion of a system's operational hours during which it is capable of performing its assigned functions. The Mean Time Between Failure (MTBF) and Mean Time To Repair (MTTR) are two elements of reliability that quantitatively express system availability, where availability = (MTBF) / (MTBF + MTTR) [DOLD78]. The computation of network availability uses the same methods. To increase availability, the MTBF must be increased and the MTTR decreased. MTBF is increased by redundant system components or by the use of proven reliable components. The MTTR can be decreased by the use of comprehensive diagnostic features for fault isolation and trouble shooting.

EXAMPLE:

The availability of the Off-Line Production Subsystem is not critical to analyst forecasting functions and as such requires an availability of only 70%.

The network must support the communications between the Analyst Subsystem and the Database Subsystem. This capability can only be unavailable one work day out of a reporting month. This represents approximately 97% availability.

The impact of the loss of each subsystem on overall system operations is explained and an availability value given for each. The ability of the system to operate in a degraded mode without network communications should be discussed. By describing system availability, the vendor is better able to understand how those requirements apply to the network.

Anticipated System Growth

System growth describes anticipated organizational, functional, or physical changes. These changes may require additional interfaces with other organizations and with the data they may provide. New functions may be added and old functions deleted that influence the network proposed by a vendor. Physical changes in the location of components, the overall size of the network, and the number of components supported have a direct impact on the network proposed.

System Loading

Established system usage patterns that may impact system performance and communication demands should be identified. Peak traffic conditions that occur with the 8 AM rush of analysts reading their mail files, the 11:30 AM rush to get a final compilation in before lunch, and the final flood of forecast batch jobs at 5 PM are described and approximated. System resource intensive jobs are identified with the frequency of their occurrence.

4.2 Network Requirements

Local network requirements are categorized in terms of:

- Services
- Traffic
- Reliability
- Growth
- Maintenance.

The following subsections describe procedures that help direct the requirements specification phase of a local network procurement. A team approach that draws from in-house expertise may be appropriate. An understanding of the following subjects is helpful when analyzing the requirements of a local area network:

- The current structure of the organization's applications (data flows, personnel responsibilities, use of existing equipment) as well as the possible ways in which these applications may use a local network.
- System programing of the major devices that are to be attached to the network (e.g. large computers or workstations.)
- Hardware interface characteristics of the user equipment (e.g. how are terminals currently attached to the computers?)

The goal of the requirements analysis is to produce a set of informal documents, each covering a specific type of network requirement. These documents are not distributed to vendors in the actual solicitation, but are used as inputs to the solicitation preparation.

4.2.1 A Generic Local Network Model. The procedures described below for specifying network requirements use a simple local network model. The model is specific enough so that requirements are easily stated; yet, is general enough so that any particular local networking technology is applicable.

The procured local network may be used to connect a wide variety of user equipment, including large computers, minicomputers, terminals, word processing stations, personal workstations, and batch processing terminals. In the model, this equipment is referred to as "user devices".

The local network vendor provides equipment for attaching user devices. This equipment varies from vendor to vendor, ranging from a centralized switch to a set of small interface units sitting between the user devices and the network's transmission medium.

The term "network access unit" (NAU) is used below as a generic reference to the vendor-provided equipment that provides user devices with access to the network.

"Network services" are accessed by the user devices through vendor-provided NAUs. It is possible that a vendor provides other equipment with the network besides the NAUs; however, only equipment specifically accessed by the user devices is described in the requirements specifications. As far as user devices are concerned, all network services are accessed via the NAUs.

4.2.2 Network Service Requirements. Preparation of the Service Requirements Document consists of three major steps:

1. Draw up an informal list of all the ways in which the network can be used. Items include specific user devices, names of departments, applications, names of specific files accessed remotely.

2. Organize the information into a "service architecture" that identifies user applications. The service architecture can be patterned after the hierarchical protocol architecture developed in the International Standards Organization (ISO) Reference Model for Open Systems Interconnection ([ISOR80], described in Section 2.3).

3. Determine the specific network services required by the "lower layer" user applications, and create a "Service Requirements Document" which is referred to when writing the solicitation.

The following is a more detailed description of the tasks involved.

STEP 1 - COLLECTION OF INFORMATION

Through discussions with department administrators, in-house system programers, engineering staff, and application end-users, answers to the following questions are gathered:

1. What is the network to be used for? These can be general answers such as, "transfer files between our two computers," or very specific answers such as, "Jim needs to get at the IMPG/X1 data from his terminal." Implied in the answers to this question are the general office functions to be supported by the network, including word processing, data entry, and electronic mail. The goal is to be as complete as possible, with no concern for the form the answers take or their relationships with other answers. For each answer, determine the name of a staff member who can be contacted for a more specific response if necessary.

2. What devices are to use the network? Be as specific as possible (e.g. model numbers). Include both devices currently installed and new devices likely to be obtained. For each device, determine the name of the most knowledgeable staff member who can be approached for further information when necessary.

3. What services must the network provide? This initial list may include vague items or overly specific items For example, the list of desired network services may include, "allow terminal connections to all computers," "virtual circuits," "support BISYNC devices," "electronic mail," and "process addressability."

The creation of these lists can proceed through informal "request for comments," interviews with key personnel, or whatever method best suits the situation. Much of the information may already have been compiled during the initial requirements analysis that determined the need for a local network or for more general office automation systems (see, for example, NBS Special Publication 500-72 [BOOA80]).

STEP 2 - DEVELOPING A SERVICE ARCHITECTURE

The goal is to determine an overall structure for the various applications identified in Step 1 and to distinguish those applications that directly use the services of the network from those that indirectly use these services.

Looping through the following 3 steps helps determine of the set of network services:

A. Develop a general service architecture that describes the overall service that the network provides to the collection of user applications.

B. Based upon the general service architecture, determine an allocation of function between each user device and the NAU to which it attaches.

C. Review the service architecture for technical feasibility, basing the review on a general knowledge of current local network technology.

Repetitions through this loop ideally result in a statement of the specific services that the user devices require from the network. Unfortunately, different user applications require different network services. To rectify this problem, arrange the possible network services into a linear hierarchy of "layers", where each layer uses services provided by lower layers and offers services to higher layers. The procured network should then provide the minimum level of service in this hierarchy meeting the needs of all applications.

Step 2 is a complex task. The process is one of refinement, starting from the data collected in Step 1 and proceeding several times through the following list of procedures:

1. Model the attachment of each user device to a network access unit in terms of its interface. If the interface is known (e.g. RS 232), the task is easier. However, in many cases a variety of attachment strategies are possible, where the decision among them is made on performance criteria rather than service criteria. Consequently, the actual details of this interface can be left unspecified at this time.

2. From the "current service architecture", determine the applications to be implemented in the user device. Determine which of these user applications directly use the network and which are "high level". For example, three user applications on the list may be "Management Information Service (MIS)," "transfer of files," and "access to the personnel records." It could then be determined that the MIS application requires access to the personnel records, which are accessed by transfers of personnel files. Hence in this example, the only application of the three that uses the network "directly" is the program that transfers personnel files.

3. By examining the structure of the applications within each device refine the current service architecture, which then includes the following:

1. A more specific description of the applications to be implemented in the user devices,

2. An arrangement of the user applications into a structured description based on the concept of "higher layer" applications using services provided by "lower layer" applications,

3. A more specific description of the services required from the local network. For example, it may be specified that the network allow for the opening of "connections" between any two user devices, with error control provided by the network. Alternatively, it may be specified that the network provide some higher layer functions such as electronic mail, or that it need only provide lower layer data transport with minimal overhead.

4. As the final step in this iterative loop, measure the interim service architecture against three yardsticks:

a. Are the identified service requirements placed on the local network likely to be satisfied by a real network currently on the market? If not, is it reasonable to assume a custom network could be procured that meets these requirements? The answers to these questions require analysis based on the review of local network technology given in Section 2.3.

b. Does the service architecture make sense for the individual applications and devices? Feedback from the staff personnel identified in Step 1 is required to ensure that the identified network service requirements meet the needs of network users.

c. Are there FIPS network standards [NBSL81] and/or guidelines in existence that apply to the layered service architecture? If so, incorporate the appropriate standards into the requirements document by reference.

The result of these reviews is a list of problems with the current draft of the service architecture, in preparation for the next pass through the loop.

Several iterations through the loop results in a cohesive system architecture The goal is a clear delineation of responsibility, distinguishing functions of the network from applications provided outside the network.

STEP 3 - SPECIFYING NETWORK SERVICE REQUIREMENTS

The final step in determining network service requirements consists of "formalizing" the results of the earlier analysis into a "Service Requirements Document". This document serves as input to the writing of the solicitation, and includes:

1. The list of FIPS network standards that are applicable to the services required [NBSL81].

2. A high-level description of the basic applications that use the network such as data entry into a set of accounting programs, program development work on several large computers and others.

3. A description of the "immediate users" of the network such as asynchronous terminals, a particular Database Management System, or user-developed file transfer programs.

4. A description of the actual services required from the network access units. The following are examples of such service requirements:

   1. Each device attached to the network shall be able to open multiple connections to other devices over which data is to be transferred reliably, with the network providing sequence and error control.

   2. The network shall ensure that no user device is capable of flooding the network in a manner which precludes the transfer of data between other devices.

   3. The network shall maintain a log of each individual's usage, and shall provide usage reports to (and only to) authorized personnel.

o A list of specific devices that are to be attached to the network. If specific interface requirements have been identified, these should also be included. Equipment configuration diagrams giving physical locations are also useful.

o Finally, a list of special requirements covering the physical requirements on the network (e.g. "must span a 5 building campus covering 15 acres"), monitoring and control requirements, and other general items that deal with the network as a whole rather than with the service provided by the network access units. A complete list of these requirements is not necessary at this time, but will emerge gradually throughout the entire requirements analysis.

4.2.3 Network Traffic Requirements. The goal of the traffic analysis task is the preparation of a "Traffic Requirements Document". This document includes all requirements related to the network's handling of the anticipated traffic volume, including throughput, delay and connection concurrency requirements. Requirements must be placed on individual network access units as well as on the network as a whole.

The specification of network traffic requirements does not begin until the service requirements have been identified. The analysis of required services included an identification of the applications that directly use the services of the network and of the user devices that attach to the network.

The following suggested steps give procedures that help identify the traffic characteristics of the installation's applications.

STEP 1 - COLLECTION OF INFORMATION

Gather the information which was generated from the service requirements task:

- A list of user devices (computers, terminals, workstations, etc.) that access the network directly by attaching to the vendor's network access units. This list is specific with repetitions if more than one device of a given type is to access the network.
- A list of those applications that directly access the services provided by the network. These applications are implemented within the user's devices.

These two lists include the names of staff personnel who can be contacted for additional information about a particular list item.

STEP 2 - DETERMINATION OF TYPICAL DATA UNIT SIZES

Each application using the network involves the transfer of data in individual "data units". The typical size of these data units is a basic characteristic of the application; their size estimates are an important part of the traffic requirements analysis.

For example, a "dumb" terminal used for interactive access to a large computer transfers individual characters one at a time. Hence the "data unit" for such interactive terminal applications is one character. In contrast, a smart word processing workstation that transfers whole pages between its local storage and that of a large computer would have a characteristic data unit of one page, which could be on the order of 2000 bytes.

In Step 2 the characteristic data unit size for each application that uses the network was considered. This determination takes the following into consideration:

1. Some applications are asymmetric, with data transfers in one direction having different patterns from those in the other direction. If possible, such situations are identified during the determination of an application's characteristic data unit size.

2. Some applications transfer large quantities of data in a stream - for example, a transfer of a large file. Since the network actually transfers data in discrete pieces, the characteristic data unit size for such applications can be given as a typical maximum transfer size for a local network, say 512 bytes

3. Many applications involve the exchanges of control information followed by data. To estimate the size of the application's characteristic data unit, concentrate on the information exchanges that dominate the traffic volume. If a particular application seems to involve exchanges following two distinct patterns, then explicitly identify two characteristic data unit sizes for that application.

The output from Step 2 is a table of applications with their characteristic data unit sizes.

STEP 3 - DETERMINATION OF DEVICE-TO-DEVICE TRANSFERS

Each user device that attaches to the network may require communication with several other devices, and may be participating in several applications simultaneously. Develop a simplified model of the typical traffic patterns between each pair of user devices.

For each pair of devices, list the applications that require exchange of data. For example, data transfers between a personal workstation and a large computer arise from word processing, accounting, data entry, interactive program development, or bulk file transfers. For each application, enter the average and peak number of individual data transfers per minute and identify special delay requirements.

In general, local networks have extremely low delay, and rarely will an application have delay requirements that cannot be met by the network. Only extreme delay requirements are explicitly identified here. It is necessary to make up two such lists for each device pair, separately covering the data transfers in the two directions.

Device A to Device B

Application Name	Minimum Delay (optional)	Data Unit Transfers/Minute Average	Peak
Application 1		XX/min	XX/min
Application 2		XX/min	XX/min
Application 3	XX msec	XX/min	XX/min

Device B to Device A

Application Name	Minimum Delay (optional)	Data Unit Transfers/Minute Average	Peak
Application 1		XX/min	XX/min
Application 2	XX msec	XX/min	XX/min
Application 3		XX/min	XX/min

STEP 4 - DETERMINATION OF INDIVIDUAL DEVICE TRAFFIC

Using the tables of device-pair traffic generated in Step 3, determine the characteristics of each device's traffic. The primary goal is the development of throughput requirements for each device's network access unit.

There are two ways to expressed a devices traffic:

1. Bytes/second (not worrying about how the bytes are packaged into separate data units), and

2. Data units/second (not worrying about the number of bytes in each separately transferred data unit).

In addition, it is necessary to separately consider data transfers into each device and out of each device, with average and peak values for both. Thus for each attached device, a table is created with 8 entries:

Device Name

Input	Data Unit Transfers/second		Bytes/second	
	Average	Peak	Average	Peak
	XX	XX	XX	XX

Output	Data Unit Transfers/second		Bytes/second	
	Average	Peak	Average	Peak
	XX	XX	XX	XX

The following paragraphs describe how to determine the values of the table entries for a particular user device.

To start, gather all device-pair traffic tables involving the device in question (from Step 3). To calculate the average number of separate data unit transfers that the device generates as output, analyze the list of applications in which the device participates. From this list of applications, estimate an "average workload" for the device; i.e. what applications are typically running most often, and how many are concurrently generating data units that must be transferred across the network to some other device. From this analysis estimate the average number of data units of various sizes that generate network transmissions.

From the list of each application's characteristic data unit size (determined in Step 2), estimate the total number of bytes that this device generates on the average each second. This estimate takes into consideration applications that typically generate data units concurrently.

A similar determination is done to calculate the average values for the device's network input, both in terms of data units per second and bytes per second. This again involves examining the device-pair tables from Step 3.

To estimate peak values for both input and output, consider the peak values for the individual applications and the maximum number of simultaneous applications generating and consuming network traffic at the given device. The peak values for the individual applications are overestimated by adding the peak value for the each device. This figure is adjusted downwards to obtain a more realistic value unless the applications involved require the network to handle simultaneous peak traffic loads of several applications. A downward adjustment to .75 of the sum of the application peak values generally gives a more realistic peak value for the device.

STEP 5 - SPECIFICATION OF NAU TRAFFIC REQUIREMENTS

After completion of Step 4, individual traffic requirements of various user devices are cataloged. These are placed into a form useful for a network procurement. The goal in Step 5 is to identify traffic requirements that can be placed on the vendor-provided Network Access Units. The specification of these requirements proceeds as follows:

1. Analyze the individual traffic characteristics of each user device from Step 4 to categorize the devices according to their network traffic requirements. For example, some devices (e.g. terminals) generate and consume small data units at a low rate. Others (e.g. large timeshared computers accessed by interactive terminals) generate and consume small data units at a high rate. And, backend computers, accessed by other computers and peripherals, generate and consume large data units at a moderate rate.

2. For each class of user device, specify the requirements that a Network Access Unit has to satisfy to be used by all devices within the class. These requirements are placed in a table identical in form to the tables generated in Step 4 for the user devices. The peak entries are derived by considering the maximum of the peak values for the various user devices that are serviced by an NAU of this type.

The end result of this analysis is the identification of a set of NAU types. Each type of NAU is given a set of traffic requirements, which if satisfied suffices to service the needs of a class of user devices.

For example, the traffic requirements in the actual solicitation is stated in the following form: "The procured network shall support the data transfers of two classes of user device, with the following traffic characteristics:

CLASS A (average and peak traffic generated and consumed by user devices of Class A)

CLASS B (average and peak traffic generated and consumed by user devices of Class B)

In some cases it is not possible to identify a small group of well-defined classes of user device. In such a situation, the installation procures a local network with a single NAU type capable of supporting all user devices. Such an NAU generally has more capacity than necessary for many user devices. Alternatively, special-purpose NAUs could be procured and tailored to the requirements of individual user devices. Such a procurement involves building a custom network.

STEP 6 - SPECIFICATION OF AGGREGATE NETWORK TRAFFIC

Finally, the average and peak loading of the network as a whole is specified. This is derived from the traffic characteristics of the individual devices as determined in Step 4.

Average aggregate traffic volume (in data units/minute and in bytes/sec) is computed by summing the average traffic output of each user device attached to the network. The peak aggregate traffic volume is also computed by summing the peak output values for the various devices. Often, the sum overstate the actual peak load that the network must handle since it assumes that all attached devices are simultaneously generating their peak volume. The actual peak aggregate traffic volume is a fraction (greater than one-half) of the value obtained by summing the individual peaks. The actual value of this fraction depends on:

1. The number of user devices

2. The critical nature of the applications

3. A rough estimate of the likelihood that various devices simultaneously generate their peak volumes of network traffic

These items are considered in determining the weighting factor for each term of the sum of peak output values. The weighting factors are multiplied against each term and the weighted terms are then summed, producing the peak aggregate traffic volume stated as a requirement that the procured network must satisfy.

Any special delay requirements noted in Step 3 are included as requirements on the network at this point.

<u>4.2.4 Network Reliability Requirements.</u> Two separate requirements are covered by "reliability":

1. Equipment availability requirements, involving the reliability of the network as a whole and of its constituent pieces (e.g. MTBF and MTTR of the transmission media, network access units, and switches).

2. Service quality requirements, involving the probabilities of service degradation during normal network operation (e.g. bit error rates, abnormal connection terminations, etc.).

An analysis of the intended network applications is necessary before either network availability or service quality requirements are determined. The step-by-step approach outlined below begins with a review of the user's applications, drawing from the work already done in the Service Requirements Document. After this review, detailed specification of reliability requirements begins.

STEP 1 - REVIEW OF APPLICATIONS' SERVICE REQUIREMENTS

During the preparation of the Service Requirements Document a rough "service architecture" was prepared. The service architecture identified applications that are direct users of the network, and identified services that provide support to those applications. The services fit within the layered protocol structure of the ISO Reference Model [ISOR80], and include:

- A simple "bit transfer" service, in which the network acts as a "bit-pipe" providing no error, sequence, or flow control at the physical and data link protocol layers.

- "Connectionless" service, in which the network ensures that a received data unit contains no errors, but does not guarantee the delivery of every datagram submitted for transmission, as at the data link, network, transport, and session layers.

- "Connection-oriented" service, in which the network provides for the establishment of virtual connections for data with the network providing error, sequence, and flow control, at the network, transport, and session layers.

- Higher layer services that fit at the presentation and application layers.

During Step 1 of this task, the network's intended applications and the services provided are analyzed. A list answering questions regarding the applications is produced:

1. How critical is the application?

2. Can the application be accomplished in the face of periodic loss of access to the network?

3. What specific network service(s) does the application require?

4. Can the application processing be done by any of a number of devices, or is one particular device necessary?

5. Can network-induced errors be recovered from easily?

Those applications intolerant of network outages, and/or incapable of recovering from network-induced errors, are marked for special consideration in the steps that follow.

STEP 2 - SPECIFICATION OF NETWORK AVAILABILITY REQUIREMENTS

A set of critical applications which must be "guaranteed" access to the network is extracted from the list developed in step 1. Since absolute guarantees are never possible, the solicitation specifies values for "Mean Time Between Failure" and "Mean Time To Repair" based on the needs of the most critical applications.

MTBF and MTTR values are specified for the network as a whole, for the individual network access units, and for any specialized equipment provided by the network vendor. These values are derived from the following considerations:

1. The network access units need not be more reliable than the various devices attach to them, provided that the failure of a network access unit does not disable the entire network.

2. The network as a whole needs to be more reliable than the user devices or the individual network access units if critical applications exist that require communications (i.e. stand-alone processing will not suffice).

3. The MTBF and MTTR figures are reasonable and within reach of current technology. If necessary, the network is configured with redundancy that meets availability requirements.

For each user device, the applications in which it participates is examined to determine the MTBF and MTTR values for its associated NAU. The solicitation includes as the MTBF/MTTR requirements for a specific NAU class the strictest MTBF and MTTR values derived.

The following list is the product of this analysis:

	MTBF	MTTR
Network	XX	XX
Class 1 NAU	XX	XX
Class 2 NAU	XX	XX
.		
.		

STEP 3 - SPECIFICATION OF SERVICE QUALITY REQUIREMENTS

The form that a service quality requirement takes is specific to the service under consideration. For example, "bit error rate" is an appropriate measure of the reliability of a network-provided "bit pipe" service that performs no error control, whereas the concept of bit error rate is essentially meaningless for a connection-oriented service that provides error-free transmission. The following table lists typical service quality metrics for the various possible network services:

Bit-pipe service

Bit error rate (e.g. one bit in a billion is in error on average)

Datagram service

Datagram error rate (e.g. one datagram in a million is not delivered)

Connection service

Blocking rate (the percentage of connection attempts that fail due to network saturation)

Abortion rate (the probability that a connection terminates abnormally due to network causes)

Detected error rate (the probability of errors detected by the network but not recovered from, with user notification)

Residual error rate (the probability of undetected errors passed on by the network to the user)

The specification of reliability requirements dealing with service quality treats each network service separately.

For each service:

1. List the applications that use the service

2. For each application, note its ability to tolerate degradations in the basic service provided (e.g. can it easily recover from network-induced errors?)

3. Determine the service quality requirements necessary to support the needs of the critical applications

Thus the output of this step is a table of the following form:

Service 1	quality requirement 1
	quality requirement 2
	.
	.
	quality requirement n
Service 2	quality requirement 1
.	.
.	.

This table is incorporated into the Reliability Requirements Document along with the availability figures generated in Step 2.

4.2.5 Network Growth Requirements. There are several ways that an installed local network grows.

STEP 1 - ANALYSIS OF EXISTING APPLICATIONS

Various applications have already been identified as users of the network (in the "Service Requirements Document" that was prepared earlier). The networking requirements of each of these applications may grow in several ways:

1. Additional traffic may be generated through devices that already participate in the application (Type 1 growth).

2. New devices may be added to the network in the future which participate in the application (Type 2 growth).

3. Other existing devices already on the network may be brought into the application (Type 3 growth).

Each type requires additional network traffic handling capacity. The first and third type in addition require Network Access Units capable of expansion to handle higher traffic volumes.

For each identified application, a list is drawn up that gives the likelihood of growth of the various types. Probabilities can be stated in terms of "high, medium, low". Thus the table has the following form:

Application X

Type of Growth	Probability of Occurrence
1	(high, medium, or low)
2	(high, medium, or low)
3	(high, medium, or low)

The second column is expanded into multiple columns giving probabilities of occurrence within specific time frames (e.g. probability of occurrence within 1 year, 3 years, etc.).

STEP 2 - IDENTIFICATION OF POSSIBLE NEW APPLICATIONS

In addition to traffic growth, new applications may be required in the future. New applications fall into two categories depending on whether or not new services are required from the network. For example, adding a voice application to an existing network requires a new service from the network. Downline loading of user devices does not require a new new service.

A lists all applications that may be added to the network identifies the probability with which the addition is likely to occur and whether any new network services are anticipated. The output of Step 2 has the following form:

New Applications	Probability of Occurrence	Description of New Network Services
Application 1	(high, med, low)	...
Application 2	(high, med, low)	...
.		
.		

If possible, the table is augmented with information describing the traffic requirements of the applications in a form similar to that given in the Service Requirements Document.

STEP 3 - SPECIFICATION OF A TOTAL GROWTH PROFILE

From the analysis performed in Steps 1 and 2, the various types of growth for the network are assigned occurrence probabilities. These probabilities are incorporated into a "Growth Requirements Document" that covers requirements for the network as a whole and for the individual network access units.

For the network as a whole, the following possible types of growth are discussed:

1. Addition of new network access units - include estimates on the number of new units anticipated to be necessary, for example, within 2 years, 3 years, and 5 years.

2. Increase in aggregate traffic volume - include quantification of anticipated bytes/sec (and data units/sec) at 2 years, 3 years, and 5 years.

3. Addition of new network services - included a description new applications and new network services anticipated within 2 years, 3 years, and 5 years.

Growth requirements may also be placed on individual network access units. Vendors may meet these requirements by either upgrading existing access units or by replacing them with more capable units. These requirements can be stated in terms of the "classes" of NAUs identified in the Traffic Requirements Document. For each class of NAU, the following growth issues are discussed:

1. Expansion of NAU throughput capacity - include quantification and specific time frames when expansion is anticipated.

2. Increase in the number of simultaneous connections handled by a single NAU.

3. Addition of new network services - include a description of new applications to be supported and new network services required; include with specific time frames.

The requirements identified in this growth analysis are incorporated into the solicitation. For example, vendors may be asked if their network meets the following growth patterns: "Two years after the initial installation, 50 additional Class 1 NAUs must be added to the network, and the total aggregate traffic volume will increase to 500 K bytes/sec. Within 3 years the network must provide encrypted connections, with network-provided key generation and distribution."

4.2.6 Network Maintenance Requirements. The goal of maintenance requirements analysis is to identify technical issues that bear on the questions of maintenance responsibilities. The "Maintenance Requirements Document" identifies in-house and vendor or third party capabilities used to support network hardware and software.

STEP 1 - ANALYZE IN-HOUSE TECHNICAL SUPPORT

A list of in-house technical capabilities is developed for the following areas:

- Digital and analog hardware.

- User-device systems programing.

- Microprocessor software.

Systems programing expertise is desirable if integration of the network services with the user devices can be staff responsibility. In this case, maintenance of network interface software within user devices is also a staff responsibility.

If the staff possesses microprocessor software expertise, the source code for the NAU could be obtained as part of the procurement. This would allow software upgrades and maintenance to be an in-house responsibility for the entire network. In addition, support tools such as hardware monitoring equipment or microprocessor development systems that already exist can be obtained.

STEP 2 - ANALYSIS OF NETWORK MAINTENANCE REQUIREMENTS

Determine the maintenance needs of the actual system to be procured. This is accomplished by a review of the system requirements developed in the Service Requirements Document and the Traffic Requirements Document.

The following issues are addressed during this analysis:

- How complex is the software likely to be within the network access units and other vendor-provided equipment? This depends on the level of service required from the network.

- How complex is the hardware technology likely to be in the procured system? This depends on the data rates supported and the services provided.

- How complex is the network interface hardware and software in the various user devices?

Informal discussions of these issues with key staff technical personnel are necessary before in-house vs. vendor or third party network maintenance is decided.

STEP 3 - SPECIFICATION OF NETWORK MAINTENANCE REQUIREMENTS

Answers to the following questions provide the basis for specifications of network maintenance requirements:

1. Should the vendor be asked to provide full maintenance of the installed hardware or can some of this responsibility be taken over by the staff?

2. Should the vendor be asked to provide monitoring equipment used by the technical staff?

3. Who is to be responsible for development and maintenance of interface software?

4. Are user modifications to the vendor-provided software required in the network devices? Should the vendor be required to provide source code for the network devices?

5. Are maintenance contracts with contractors other than the original vendor necessary or desirable?

The final version of the maintenance requirement document depends on a cost analysis as well as the technical analysis performed.

4.3 Network Procurement Requirements

The subsections that follow discuss the network life-cycle, government furnished equipment, documentation and modifications to the network.

<u>4.3.1 Network Life-Cycle.</u> Phases in the network life cycle include:

1. Incremental delivery schedule.

2. Installation.

3. Acceptance.

4. Operation, including maintenance and training.

Government and vendor responsibilities during these phases should be specified in the requirements.

The schedule for network procurement should identify equipment and services, when deliveries must be made, and in what form. A number of installation-specific issues may impose requirements on the performance schedule, such as building construction, priority of services to be introduced, personnel assignments, delivery and acquisition procedures for non-network equipment. The schedule defines the dates and quantification for:

Services:

It may be desirable to introduce services in phases according to the priorities of groups. Test beds exhibiting initial services may first be installed to demonstrate basic communication capabilities, followed by higher level and more extensive services in phases according to their priority and the agency's ability to make use of them.

User Groups:

Full or partial network capability may be introduced in phases by organizational groups, such as:

Group Y shall have Capability A on XX date,

Group Z shall have Capabilities B and C on YY date.

Equipment:

Capabilities may be introduced when the installation has certain types of devices in operation. If, for example, a group of workstations, computers, and printers are operational on XX date, then communication to those devices must be available on YY date.

Performance:

Performance classes may be introduced in a phased schedule. Depending on the traffic load expected over varying time periods, increments in network performance may be introduced accordingly. Upgrades from lower to upper performance classes may be desirable or at least acceptable.

A variety of component installations affect the network delivery schedule.

Transmission media layout and installation is the first step. The media are specified and installed according to applicable Government regulations; it may be advisable that it be installed according to growth projections as well as initial capability requirements. Once installed, the transmission media is tested for conformance to electrical and physical specifications. Immunity from improper termination must also be checked as well as the performance of standby equipment.

Once the medium is checked, the other network components are integrated and checked on a one-by-one basis, followed by network access unit (NAU) to user device tests. As those are passed, network subsystem integration tests are run to verify operational capabilities of the network. Network to system integration represents the final installation phase.

Once the vendor has verified the operation of the network formal acceptance testing begins. A comprehensive acceptance test program consists of a combination of the following verification methods:

1. Visual examination.

2. Engineering diagnostic tests.

3. Functional tests.

4. Demonstrations.

Testing requirements in each area are defined in the solicitation.

The visual examination is a configuration audit to verify workmanship and that electrical, mechanical, and safety requirements are met. It also confirms the hardware inventory.

Engineering diagnostic tests primarily consist of running a vendor supplied diagnostic package. This package should exercises all device functions and modes of operation.

Demonstrations are often software programs that emulate aspects of an operational network. These programs exercise interfaces to the network and demonstrate synergistic interaction of components.

Once the network is accepted, either entirely or in phases as specified, network operation begins. At this point the federal agency may require additional vendor support or use internal personnel for maintenance and user training.

Maintenance of the network takes two forms: preventive, which is performed periodically to check equipment function and performance, and curative, which is performed when problems in equipment or software functions are detected by agency personnel. Requirements define:

- Who is to perform the maintenance functions, including any restrictions on numbers of personnel that can be on-site and their security clearance levels, if applicable.
- The availability of user devices in terms of amount of time and periods for problem trouble-shooting.
- Where maintenance is to be performed: at the vendor facility or at the agency installation.
- Limitations on the time to correct or respond to notification of error.
- Any special working conditions to which vendor personnel must adhere.
- Special maintenance contract considerations, such as: "Maintenance must be performed at no charge to the Government for a specified period of time".

The Federal agency may require that its personnel be trained for use of network services and/or maintenance of the network equipment and software. The requirements specify the level of training in terms of:

- Hours per day.
- Personnel to be trained, characterizing their experience level and training objectives.
- Time when training takes place: pre-installation, during installation, post-installation.

Location for training is specified, in addition to expected duration of classes and availability of system equipment as training aids.

4.3.2 Government Furnished Equipment (GFE). If the requesting agency has existing equipment that is required as part of the network, descriptions of this equipment are included in the solicitation document. The procuring agency states when and what number of GFE units shall be made available to the vendor. These include existing media systems, media access units, and test equipment

Descriptions of how this equipment is currently used by the procuring agency helps the vendor determine if the GFE units can be integrated with their products.

4.3.3 Documentation. The documentation requirements are for an off-the-shelf network procurement. Documentation for a custom network requires detailed hardware and software design documentation in addition to off-the-shelf products. The following documentation is required for custom networks:

Interface Specifications,

General interface specifications to their various networking products, and

Interface documentation for each specific user device that must be supported by the network.

Maintenance Documents,

Hardware maintenance descriptions and schedules for each network hardware component provided by the vendor, and

Software maintenance documents that contain the general program information, program descriptions, operating environment description, and maintenance procedures.

User Manuals,

Hardware, and

Software User's Manuals containing general information about the software, its application, initiation, inputs, outputs, and error recovery as specified in FIPS PUB 38 [FIPS76].

Test Plans and Procedures for network acceptance, including:

Hardware tests and procedures, and

Software tests and procedures [FIPS76].

4.4 Solicitation Preparation Summary

The first step in preparation of the local network solicitation document is to provide descriptions of the system. These descriptions focus on:

- System applications,
- Services,
- Equipment configuration,
- Data flow,
- Operations,
- Characteristics: performance, reliability/availability, anticipated growth, loading, and
- Responsibilities.

These descriptions are producing in a set of five informal "requirements documents" used during the final writing of the solicitation.

These documents identify specific services required of the network. Analysis of each application is performed to determine a generalized service specification for network-wide use. Once specified, performance requirements are determined by examining average and peak data rates. Then specifications for reliability, growth, and maintenance requirements are produced.

Clear specifications of incremental delivery schedules and vendor responsibilities throughout the network life-cycle are made; these specifications include: installation, acceptance, operation, maintenance, and training.

Other procurement considerations are important. Documentation for users and operations personnel is required, integration with existing GFE may be required and mechanisms that allow for network modifications and growth may be required.

5. EVALUATION AND SELECTION PROCESS

Evaluation of technical proposals is a complex process; but, it can be simplified using quantitative methods. The method makes use of the importance and priority of each requirement, and the capabilities of the vendor to satisfy the requirements.

Prior to issuing the solicitation, a series of weighting factors corresponding to each requirement are developed. Qualification values derived from these weights are summed producing a numerical score representing the vendor's total proposal performance.

5.1 Quantitative Development

A quantitative method is presented in the four subsections that follow. The first section describes the method; the second section presents guidelines for developing weighting values; and, the third describes evaluation of vendor qualifications. A discussion of the vendor's score follows as the final section.

5.1.1 The Quantitative Evaluation Method. The quantitative method has four steps. The first is performed when the requirements themselves are finalized. The second and third are performed for each vendor; the fourth step produces the highest scoring bid.

Step One

Step One of the method takes place prior to issuance of the solicitation after all requirements have been specified and evaluated in terms of priority and relative importance. Each requirement is given a weighting factor, Ra. The purpose of this step is to develop the set of requirement weighting factors, R1 through Rn, where "n" is the number of individual requirements. The values Ra may be any fraction between 0 and 1 expressed in decimal form.

Once the solicitation is issued, the R1 through Rn factors become fixed.

Step Two

This step starts when all of the bids are received. The purpose is to develop a set of vendor qualification ratings for each bid, where each rating corresponds to each requirement specified in the solicitation. Each proposal is given a set of ratings, Qa.

For each vendor proposal, develop the set of vendor qualification ratings Q1 through Qn, where "n" is the number of requirements specified in the solicitation. The values may be any integer or fractional value between 0 and 10, expressed in decimal form.

The Q values must be computed fairly, where all proposals are evaluated simultaneously for each individual requirement. This allows evaluation of the relationship between proposals on a requirement by requirement basis.

Step Three

The third step develops a single "score" for each proposal. Each vendor qualification rating is multiplied by the respective requirement weighting factor in order to place the proper emphasis on the vendor's solution to that requirement. The resulting values are summed to produce the score for that proposal.

Evaluate the following equation for each bid:

$$(R1 * Q1) + (R2 * Q2) + \ldots + (Rn * Qn) = Sa$$

where Sa represents the score of bid "a".

This step produces S1 through Sm, where "m" is the total number of bids received.

Step Four

The final step chooses the bid with the highest score and makes the award based on that score considering a separate evaluation of the cost. Cost evaluation is outside the scope of these guidelines.

5.1.2 Requirement Weighting Factors. Each requirement weighting factor is expressed as a decimal fraction between 0 and 1; high priority requirements receive a value close to 1. The following assessment results in a weighted average:

1. Determine the relationship of the requirement to other network requirements. Does it represent a service or mandatory item or is it a low priority item?

2. Determine the flexibility and range of the function and performance of the requirement.

3. Determine if the requirement fits within a fuzzy area between network and system function.

4. The ease with which this process is accomplished depends on a concise statement for each requirement and on its suitability to the system environment and application. Assessment of a proposed solution also requires a concise statement of requirements. An additional check on that assessment requires inclusion of requirements that demand a:

   1. Detailed description of the network in global terms, with the mechanisms for total system solution (Section Four).

   2. Specification of equipment and user interfaces, detailing commands, parameter settings, interface lines, etc.

   3. Diagrams of information flow and network structure.

The purpose is to ensure that vendor proposals are also evaluated based on global considerations with believable solutions.

<u>5.1.3 Vendor Qualification Ratings.</u> Each vendor proposal is evaluated on a requirement by requirement basis, with careful consideration of the solution presented to determine how well the requirement has been met. Each requirement solution is given a rating as any decimal value between 0 and 10. A generalized rating scale is as follows:

0 No. Does not meet the requirement.

0.1 - 3.0 Barely. Meets at least a portion of the requirement, but the solution is barely within or above the limitations of vendor capability.

3.1 - 6.0 Adequate. Meets the requirement and is easily within the vendor capability.

6.1 - 9.9 Exceeds. More than meets the requirement by providing additional capability and/or performance.

In developing the vendor qualification rating for a single requirement, the following assessments is made, using an average or weighted average:

1. Does the solution meet the requirement, using the generalized rating scale presented above.

2. Examine the risk involved in developing that solution. Higher risk items receive a lower rating, with lower risk solutions given a high rating. Risk assessments might include:

   a. Is the requirement met by using "off-the-shelf" products? If so, are the products well-known and trusted or have customer references been provided and checked by the agency. Customer references may be required by the solicitation or may be the result of technical questions to the vendor.

   b. Determine the extent of vendor custom-development necessary to meet the requirement, including software and/or hardware. In particular, the evaluator must determine:

      i. If new technology is required or if modifications to existing technology are proposed. New technology imparts higher risk.

      ii. If the size of the development effort in terms of labor, schedule, and material is realistic for the requirements to which it applies.

      iii. If the timeframe for development fits with the agency schedule and any phased introduction of service.

      iv. If the timeframe provides adequate time for test, integration, and acceptance procedures and is still able to meet agency schedule.

   c. If the solution includes products that are not yet available from the vendor, will the products be realistically available within the required timeframe? Does the product development schedule appear overly optimistic or is there enough time to allow for slips in the development and still meet agency needs?

3. If the solution does not specifically meet the requirement has a realistic alternative been presented, and if so, can it be used by the agency in a similar or improved fashion? In general, evaluators may consider assigning a lower ranking to an alternative than to a solution which truly meets the requirement, unless the alternative is judged as a better approach.

4. Does the solution comply with national and international standards?

5. Is the solution feasible in light of the technology and analysis presented in Sections Two and Three, and is it cost-effective according to agency needs?

At points in the evaluation process, it may become evident that technical questions must be asked of vendors who have submitted bids in order that the evaluators receive additional information. Procedures for submission of questions is outside the scope of these guidelines; the questions may include the following:

1. Clarification or resolution of information presented.

2. Additional and/or more detailed descriptive information in the presentation of solutions to requirements.

3. References to customers who are presently making use of products and/or services proposed.

4. Definition of production schedules with current milestones.

5. Descriptions of testing methods and detailed procedures used, to determine if products are fully tested before being shipped.

As the qualification ratings are assigned, the evaluators must remain cognizant of the global view of the system, its environment, and its users. The vendor must present a solution which has been adequately described, such that the evaluators can fully understand its operation and therefore make both the global and detailed assessments necessary. The ease with which the solution fits into the entire system must be readily apparent.

5.2 An Example Evaluation Process

This section presents a simple and brief example of the quantitative evaluation method for a simple local network. Six requirements have been specified:

1. The local network shall provide asynchronous serial interfaces for up to 100 terminal devices.

2. The local network shall support terminal speeds of 9600 bits per second.

3. The local network shall provide five computer interfaces for the computers in current operation at the Government facility.

4. The local network shall support terminal to computer communications with full interconnectivity.

5. The local network shall support computer to computer communications with full interconnectivity.

6. The local network shall support a single computer-to-computer communication link at speeds of at least 19.2K bits per second.

Before the solicitation was issued, an assessment of the importance of each requirement was made and weights were assigned.

1. Requirement #1 was given a weighting factor of 0.75, as 100 terminals have already been purchased and must be connected by the network as soon as possible.

2. Requirement #2 received 0.6; users will not tolerate slow response, but will actually be satisfied if 4800 bps is the implemented speed.

3. Five computers are presently in operation on site, and all five must be part of the network. The assigned weight for requirement #3 is 0.9.

4. All users must be capable of accessing all computers when they are available, yielding a weighting factor for requirement #4 of 0.85.

5. Computers must be able to inter-communicate eventually, but the capability may be delayed based upon the actual agency application expansion pattern. The factor given to requirement #5 is 0.45.

6. Considering that the computer-to-computer interconnectivity is of low importance, the speed required has a lower priority, 0.3.

Two bids were received, A and B. Evaluations of their solutions are as follows:

1. Both proposals presented a solution that easily handles 100 terminals. In addition, A presented a capability for up to 200 terminals, receiving a 7.5 rating. B received 6.0, since 100 was a realistic maximum for the capability.

2. B presented a capability for supporting all 100 terminals at 9600 bps; 6.0 is the rating. A, however, handles half the 100 terminals at 9600 bps; the other half from operate at 2400 to 4800 bps. A rates 2.5.

3. Both bids provide host computer interfaces to the network, and are capable of handling five or more. A, however, will not be able to support the type of computer interface required without a custom development effort on a tight schedule. B receives 7.0, A receives 3.5.

4. Both bids support full terminal to computer interconnectivity, with B allowing the connections to be made symbolically. A receives 6.0, B receives 7.5.

5. Proposals A and B support the full interconnectivity for transactions and connections, but proposal B requires additional commands that lower performance. A receives 6.0, B receives 3.5. This implies that speed is more important to the agency than the number of concurrent computer connections.

6. A can support the necessary speed if no more than three of the five computers communicate concurrently; A rates 5.5. B cannot support more than 9600 bps; B rates 3.5.

Given the weighting factors and qualification ratings, the score for Proposal A is as follows:

(0.75 * 7.5) + (0.6 * 2.5) + (0.9 * 3.5) + (0.85 * 6.0) + (0.45 * 6.0) + (0.3 * 5.5)
= 19.725

Similarly computed, the score for Proposal B is:

(0.75 * 6.0) + (0.6 * 6.0) + (0.9 * 7.0) + (0.85 * 7.5) + (0.45 * 3.5) + (0.3 * 3.0)
= 23.25

5.3 The Selection Process and Final Result

Final selection of the proposal that best meets the requirements of the solicitation should not be based solely upon the results of this quantitative method. Taking the proposal with the highest score is a suggestion; the final decision includes an evaluation of the cost proposal.

5.4 Evaluation and Selection Summary

Evaluation of vendor responses is simplified using quantitative methods. The importance points to consider are the priority of each requirement to the total installation and the capability presented by the vendor in satisfying the requirement.

Four basic steps are involved. Step One takes place prior to issuing the solicitation. A weighting factor is assigned to each requirement in terms of priority and relative importance. Step Two starts with receipt of all vendor proposals; a set of vendor qualification ratings corresponding to the vendor response for each requirement is developed. Step Three developments a single score for each proposal by evaluating the results using the requirement rating values and the vendor qualification ratings. Step Four chooses the proposal with the highest score.

It is important to understand that final selection should not be based solely on the results of the quantitative method; proposals may receive scores that are very close in value; cost proposals must also be evaluated and considered.

6. CONCLUSION

Selection of the local area computer network that fits the needs of the target installation requires the efforts of both technical and management personnel knowledgeable of information flow within the intended environment. Knowledge may be based upon operational experience and/or upon studies performed by internal personnel or consultants.

This guideline recommends that the agency proceed in four steps:

1. Analyze and understand contemporary local networks, putting expectations into perspective.

2. Perform analysis to determine the requirements of the installation, including services, traffic, reliability, growth, and maintenance.

3. Prepare the solicitation document, identifying not only specific network requirements but also those of the total system environment and of general network procurement.

4. Evaluate and select the local network from the proposals received, carefully using the quantitative method described.

The information and procedures presented has been based upon current local network technology with a pragmatic approach toward the problems faced by federal agencies. By studying this document and the more detailed references provided, the agency is well prepared for selection of a local network that meets its requirements and is within procurement constraints.

7. REFERENCES

[ACM65] Communications of the ACM, Vol. 8, No 5, May 1965, pp. 280-286.

[BOOA80] Booz-Allen and Hamilton, Inc., "Guidance on Requirements Analysis for Office Automation Systems", NBS Special Publication 500-72, December 1980.

[BERA80] A.A. Bergh, J.A. Copeland, R.W. Dixon, "Optical Sources for Fiber Transmission Systems", Proceedings of the IEEE, Vol. 68, No. 10, October, 1980.

[BIBK80a] K. Biba and G. Ennis, "Protocol Architecture Model Analysis". Sytek, Inc. TR-8087, 30 November, 1980.

[BIBK81b] K.J. Biba and H. Kanakia, "Local Network Architecture for Office Automation", National Bureau of Standards, ICST/LANP-81-1, May, 1981.

[BLAR80] R.P. Blanc and J.F. Heafner, National Bureau of Standards. "The NBS Program in Computer Network Protocol Standards". Proceedings of the Fifth International Conference on Computer Communication, October, 1980.

[BOGD80] D. Boggs, J. Shoch, E. Taft, and R. Metcalfe, "PUP: An Internetwork Architecture". IEEE Transactions on Communications, April 1980.

[BOLB80] Bolt Beranek and Newman, Inc. "Features of the Internetwork Protocol", National Bureau of Standards, ICST/HLNP-80-8, July, 1980.

[BOLB81] Bolt Beranek and Newman, Inc. "Specification of a Draft Message Format Standard". National Bureau of Standards, ICST/CBOS-80-2, September, 1981.

[BURJ80] John Burruss, "Features of the Transport and Session Protocols". National Bureau of Standards, ICST/HLNP-80-2, March, 1980.

[BURJ81a] John Burruss, et al., "Specification of the Transport Protocol Volume 1: Overview and Services", National Bureau of Standards, ICST/HLNP-81-11, September, 1981.

[BURJ81a] John Burruss, et al., "Specification of the Transport Protocol Volume 2: Basic Class Protocol", National Bureau of Standards, ICST/HLNP-81-12, September, 1981.

[BURJ81a] John Burruss, et al., "Specification of the Transport Protocol Volume 3: Extended Class Protocol", National Bureau of Standards, ICST/HLNP-81-13, September, 1981.

[BURJ81a] John Burruss, et al., "Specification of the Transport Protocol Volume 4: Network Interfaces", National Bureau of Standards, ICST/HLNP-81-14, September, 1981.

[BURJ81b] John Burruss, Gregory Pearson, and Thomas Blumer, "Specification of the Session Protocol". National Bureau of Standards, ICST/HLNP-81-2, March, 1981.

[BUXW80] Werner Bux, "Local-Area Subnetworks: A Performance Comparison". IBM Zurich Research Laboratory, August 19, 1980.

[CALR81a] Ross Callon, "Specification of the Internet Protocol". National Bureau of Standards, ICST/HLNP-81-6, May 1981.

[CALR81b] Ross Callon, "Specification and Analysis of Local Area Network Architecture Based on the ISO Reference Model". National Bureau of Standards, ICST/LANP-81-1, April 1981.

[CARA75] A. Bruce Carlson, "Communication Systems: An Introduction to Signals and Noise in Electrical Communication". New York: McGraw Hill, 1975, Second Edition.

[CCIT77a] Recommendation V41 "Data Transmission Over the Telephone Network: Series V Recommendations", CCITT Orange Book, VII.1, International Telecommunications Union, Geneva 1977.

[CCIT77b] Provisional Recommendation X25 "Interface Between Data Terminal Equipment (DTE) and Data Circuit-Terminating Equipment (DCE) for Terminals Operating in the Packet Mode on Public Data Networks", CCITT Orange Book, VII.2, Public Data Networks, Geneva, 1977.

[CERV78] V. Cerf and P. Kirstein, "Issues in Packet-Netwwork Interconnection". Proceedings of the IEEE, Vol. 66, NO. 11, November 1978.

[CLAD78] David D. Clark, Kenneth T. Pogran, David P. Reed, "An Introduction to Local Area Networks", Proceedings of the IEEE, Vol. 66, No. 11, November, 1978.

[CLOS80a] Samuel E. Clopper, "Features of the File Transfer Protocol (FTP) and the Data Presentation Protocol (DPP)", National Bureau of Standards, ICST/HLNP-80-6, September 1980.

[CLOS80b] Samuel E. Clopper and John E. Swanson, "Service Specification of the File Transfer Protocol (FTP) and the Data Presentation Protocol (DPP)", National Bureau of Standards, ICST/HLNP-80-12, October 1980.

[COHD78] D. Cohen, "A Protocol for Packet Switching Voice Communication". Computer Network Protocols (A. Danthine ed.), Universite de Liege, 1978.

[DAV79] Davies, Barber, Price, Solomonides, "Computer Networks and Their Protocols". Chichester: John Wiley & Sons, 1979.

[DOLD78] Dixon R. Doll, "Data Communications: Facilities, Networks, and Systems Design". New York: John Wiley & Sons, 1978.

[ENNG81a] G.B. Ennis and M.K. Graham, "Requirements Analysis of Local Area Computer Networks. Draft Report", prepared for the National Bureau of Standards by Sytek, Inc. ICST/LANP-81-3, July, 1981.

[ENNG81b] G.B. Ennis, L.G. Gardner, and M.K. Graham, "Specification of Functional Requirements for Local Area Computer Networks. Draft Report", prepared for the National Bureau of Standards by Sytek, Inc. ICST/LANP-81-4, July, 1981.

[FEDE80] "Federal Procurement Regulations 342", Second Edition. Amendments through August, 1980.

[FIPS76] FIPS PUB 38. "Guidelines for Documentation of Computer Programs and Automated Data Systems". U.S. Department of Commerce/National Bureau of Standards, February 15, 1976.

[FIPS79a] FIPS PUB 60-1. "I/O Channel Interface". U.S. Department of Commerce/National Bureau of Standards, August 27, 1979.

[FIPS79b] FIPS PUB 61. "Channel Level Power Control Interface". U.S. Department of Commerce/National Bureau of Standards, February 16, 1979.

[FORJ79] J. Forgie, "ST - A Proposed Internet Stream Protocol". Integrated Evaluation Network (IEN) 119, September 1979.

[GARL81] L.G. Gardner and M.K. Graham, "Feature Analysis of Local Area Computer Networks. Draft Report", National Bureau of Standards ICST/LANP-81-2, July, 1981.

[GITI78] I. Gitman and H. Frank, "Economic Analysis of Integrated Voice and Data Networks: A Case Study". Proceedings of the IEEE, Vol. 66, No. 11, November, 1978.

[HAFE76] E. Hafner, "Enhancing the Availability of a Loop System by Meshing", 1976 international Zurich Seminar on Digital Communications.

[HOLK76] Kenneth Dean Holberger, "An Addressable Ring Conveyor", Dept. of Computer Science, University of Illinois at Urbana-Champaign, Urbana, Illinois, January, 1976.

[IEEE78] IEEE Std 488-1978 "IEEE Standard Digital Interface for Programmable Instrumentation", Institute of Electrical and Electronics Engineers, Inc. November 30, 1978.

[IEEE76a] IEEE Std 596-1976, "IEEE Standard Parallel Highway Interface System (CAMAC)", Institute of electrical and Electronics Engineers, Inc. September 9, 1976.

[IEEE76b] IEEE Std 683-1976, "IEEE Recommended Practice for Block Transfers in CAMAC Systems", Institute of Electrical and Electronics Engineers, Inc. October 29, 1976.

[IEEE80] "Local Area Network Data Link Control", DLMAC Subcommittee of IEEE Project 802 on Local Area Network Standards, working paper, September 1980.

[ISO76] International Standard 3309, Data communication - High-level data link control procedures - Frame Structure, Ref No. ISO 3309-1976(E).

[ISO82] Data Processing - Open Systems Interconnection - Basic Reference Model, International Organization for Standardization, Draft Proposed Standard 7498, February 1982, (ISO/TC97/SC16 N890).

[KAHR78] R.E. Kahn, S.A. Gronemeyer, J. Burchfiel, R.C. Kunzelman, "Advances in Packet Radio Technology". Proceedings of the IEEE, Vol. 66, No. 11, November, 1978.

[KERI76] I.H. Kerr, G.R.A. Gomberg, W.L. Price, and C.M. Solomonides, "A Simulation Study of Routing and Flow Control Problems in a Hierarchically Connected Packet Switching Network". Proceedings of the International Computer Conference, Toronto, August, 1976.

[KLEL76] L. Kleinrock, "Queuing Systems", Volume 2, Wiley-Interscience, New York, 1976.

[LAMS80] S. Lam, "Packet Broadcast Networks - A Performance Analysis of the R-ALOHA Protocol", IEEE Transactions on Computers, Vol C-29, No. 7, July, 1980.

[LITI80] Tingye Li, "Structures, Parameters, and Transmission Properties of Optical Fibers", Proceedings of the IEEE, Vol. 68, No. 10, October 1980.

[MAFE74] E.R. Mafner, Z. Nenadal, and M. Tschanz, "A Digital Loop Communication System", IEEE Transactions on Communications, June, 1974.

[METR76] R. Metcalfe and D. Boggs, "Ethernet: Distributed Packet Switching for Local Computer Networks", Communications of the ACM, 19, 7, July 1976.

[MIT79] Experiment Plan for the Wideband Integrated Network - Supplement 1, MIT Lincoln Laboratory, December 1979.

[NBSL81] NBS Publication List 58, "Federal Information Processing Standards Publications (FIPS PUBS)". U.S. Department of Commerce, National Bureau of Standards, Institute of Computer Sciences and Technology, February, 1981.

[OHY77] Young Oh and Ming T. Liu, "Interface Design for Distrubuted Control Loop Networks", National Telemetering Conference, Los Angeles, 1977.

[PEAG80] Gregory Pearson and John Burruss, "Service Specification of Transport and Session Protocols", National Bureau of Standards, ICST/HLNP-80-2, March 1980.

[POUL78] L. Pouzin and H. Zimmermann, "A Tutorial on Protocols". Proceedings of the IEEE, Vol. 66, NO. 11, November 1978.

[POUL80] L. Pouzin, "Methods, Tools and Observations on Flow Control in Packet-Switched Data Networks". SCH 625, June 1980.

[POWC79] C. Emery Powell, "Industry Gears Up for Free-Space Communications". Optical Spectra, Vol. 13, No. 6 & 7, June, July, 1979.

[REAC75] Cecil C. Reames and Ming T. Liu, "A Loop Network for Simultaneous Transmission of Variable Length Messages", Second Annual Symposium on Computer Architecture, January, 1975.

[SCHM80] Morton I Schwartz, Paul F. Gagen, and Manuel R. Santana, "Fiber Cable Design and Characterization", Proceedings of the IEEE, Vol. 68, No. 10, October 1980.

[SHEC80] Carolyn B. Shelton and James R. Moulton, "Virtual Terminal Feature Analysis". National Bureau of Standards, ICST/HLNP-80-10, March, 1981.

[SHOJ79] J. Shoch and J. Hupp, "Performance of an Ethernet Local Network: A Preliminary Report", Proceedings of the Local Area Communications Network Symposium, Boston, May 1979.

[TOBF78] F. Tobagi, et al., "Modeling and Measurement Techniques in Packet Communication Networks". IEEE Proceedings, Vol. 66 No. 11, November, 1978.

[TOBF79] F. Tobagi and V.B. Hunt, "Performance Analysis of Carrier Sense Multiple Access with Collision Detection". Proceedings of the LACN Symposium, May, 1979.

[ZAFP73] Pitro Zafiropulo, "Reliability Optimization in Multiloop Communication Networks", IEEE Transactions on Communications, Vol. COM-21, No. 8, August, 1973.

NBS-114A (REV. 2-80)

U.S. DEPT. OF COMM. **BIBLIOGRAPHIC DATA SHEET** *(See instructions)*	1. PUBLICATION OR REPORT NO. NBS SP 500-96	2. Performing Organ. Report No.	3. Publication Date November 1982

4. TITLE AND SUBTITLE

Computer Science and Technology:

The Selection of Local Area Computer Networks

5. ~~AUTHOR(S)~~ Edited by:
Robert Rosenthal

6. PERFORMING ORGANIZATION *(If joint or other than NBS, see instructions)*

NATIONAL BUREAU OF STANDARDS
DEPARTMENT OF COMMERCE
WASHINGTON, D.C. 20234

&

Sytek, Incorporated
1153 Bordeaux Drive
Sunnyvale, CA 94086

7. Contract/Grant No.

8. Type of Report & Period Covered
Final

9. SPONSORING ORGANIZATION NAME AND COMPLETE ADDRESS *(Street, City, State, ZIP)*

Institute for Computer Sciences and Technology
National Bureau of Standards
Washington, DC 20234

10. SUPPLEMENTARY NOTES

Library of Congress Catalog Card Number: 82-600635

□ Document describes a computer program; SF-185, FIPS Software Summary, is attached.

11. ABSTRACT *(A 200-word or less factual summary of most significant information. If document includes a significant bibliography or literature survey, mention it here)*

These guidelines present features available in contemporary local area computer networks including distinctions between network applications, topology, protocol architecture and transmission media. Guidance is given to identify the installation's needs. These needs are described in terms of network reliability, traffic characterizations, expected growth, and maintenance requirements.

12. KEY WORDS *(Six to twelve entries; alphabetical order; capitalize only proper names; and separate key words by semicolons)*

Feature analysis; guidelines; local area networks; local network specification; requirements analysis.

13. AVAILABILITY

☒ Unlimited
□ For Official Distribution. Do Not Release to NTIS
☒ Order From Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402.
□ Order From National Technical Information Service (NTIS), Springfield, VA. 22161

14. NO. OF PRINTED PAGES
133

15. Price
$6.00

☆U.S. GOVERNMENT PRINTING OFFICE: 1982-380-997/2323

ANNOUNCEMENT OF NEW PUBLICATIONS ON COMPUTER SCIENCE & TECHNOLOGY

Superintendent of Documents,
Government Printing Office,
Washington, DC 20402

Dear Sir:

Please add my name to the announcement list of new publications to be issued in the series: National Bureau of Standards Special Publication 500-.

Name ____

Company ____

Address ____

City ________________ State __________ Zip Code _______

(Notification key N-503)

NBS TECHNICAL PUBLICATIONS

PERIODICALS

JOURNAL OF RESEARCH—The Journal of Research of the National Bureau of Standards reports NBS research and development in those disciplines of the physical and engineering sciences in which the Bureau is active. These include physics, chemistry, engineering, mathematics, and computer sciences. Papers cover a broad range of subjects, with major emphasis on measurement methodology and the basic technology underlying standardization. Also included from time to time are survey articles on topics closely related to the Bureau's technical and scientific programs. As a special service to subscribers each issue contains complete citations to all recent Bureau publications in both NBS and non-NBS media. Issued six times a year. Annual subscription: domestic $18; foreign $22.50. Single copy, $4.25 domestic; $5.35 foreign.

NONPERIODICALS

Monographs—Major contributions to the technical literature on various subjects related to the Bureau's scientific and technical activities.

Handbooks—Recommended codes of engineering and industrial practice (including safety codes) developed in cooperation with interested industries, professional organizations, and regulatory bodies.

Special Publications—Include proceedings of conferences sponsored by NBS, NBS annual reports, and other special publications appropriate to this grouping such as wall charts, pocket cards, and bibliographies.

Applied Mathematics Series—Mathematical tables, manuals, and studies of special interest to physicists, engineers, chemists, biologists, mathematicians, computer programmers, and others engaged in scientific and technical work.

National Standard Reference Data Series—Provides quantitative data on the physical and chemical properties of materials, compiled from the world's literature and critically evaluated. Developed under a worldwide program coordinated by NBS under the authority of the National Standard Data Act (Public Law 90-396).

NOTE: The principal publication outlet for the foregoing data is the Journal of Physical and Chemical Reference Data (JPCRD) published quarterly for NBS by the American Chemical Society (ACS) and the American Institute of Physics (AIP). Subscriptions, reprints, and supplements available from ACS, 1155 Sixteenth St., NW, Washington, DC 20056.

Building Science Series—Disseminates technical information developed at the Bureau on building materials, components, systems, and whole structures. The series presents research results, test methods, and performance criteria related to the structural and environmental functions and the durability and safety characteristics of building elements and systems.

Technical Notes—Studies or reports which are complete in themselves but restrictive in their treatment of a subject. Analogous to monographs but not so comprehensive in scope or definitive in treatment of the subject area. Often serve as a vehicle for final reports of work performed at NBS under the sponsorship of other government agencies.

Voluntary Product Standards—Developed under procedures published by the Department of Commerce in Part 10, Title 15, of the Code of Federal Regulations. The standards establish nationally recognized requirements for products, and provide all concerned interests with a basis for common understanding of the characteristics of the products. NBS administers this program as a supplement to the activities of the private sector standardizing organizations.

Consumer Information Series—Practical information, based on NBS research and experience, covering areas of interest to the consumer. Easily understandable language and illustrations provide useful background knowledge for shopping in today's technological marketplace.

Order the **above** *NBS publications from: Superintendent of Documents, Government Printing Office, Washington, DC 20402.*

Order the **following** *NBS publications—FIPS and NBSIR's—from the National Technical Information Services, Springfield, VA 22161.*

Federal Information Processing Standards Publications (FIPS PUB)—Publications in this series collectively constitute the Federal Information Processing Standards Register. The Register serves as the official source of information in the Federal Government regarding standards issued by NBS pursuant to the Federal Property and Administrative Services Act of 1949 as amended, Public Law 89-306 (79 Stat. 1127), and as implemented by Executive Order 11717 (38 FR 12315, dated May 11, 1973) and Part 6 of Title 15 CFR (Code of Federal Regulations).

NBS Interagency Reports (NBSIR)—A special series of interim or final reports on work performed by NBS for outside sponsors (both government and non-government). In general, initial distribution is handled by the sponsor; public distribution is by the National Technical Information Services, Springfield, VA 22161, in paper copy or microfiche form.