NTIA-REPORT-81-57

Performance Parameters for Digital and Analog Service Modes

M. Nesenbergs
W.J. Hartman
R.F. Linfield

U.S. DEPARTMENT OF COMMERCE
Philip M. Klutznick, Secretary

Henry Geller, Assistant Secretary
for Communications and Information

January 1981

Preface

This report is submitted in compliance with delivery requirements for the first phase of a multiphase project to develop technical performance criteria for use in the design and specification of transmission facilities for the digital defense communication system. This project was conducted for the Defense Communications Engineering Center in Reston, Virginia, by the Institute for Telecommunication Sciences in Boulder, Colorado, on Project Order Document Number DCFR 940065.

The views, opinions, and findings contained in this report are those of the authors and should not be construed as an official Defense Communications Agency policy or decision unless designated by other official documentation.

Certain commercial equipment, instruments, protocols, or materials are identified in this report to adequately specify the experimental procedure. In no case does such identification imply recommendation or endorsement by the National Telecommunications and Information Administration, nor does it imply that the material or equipment identified is necessarily the best available for the purpose.

TABLE OF CONTENTS

TABLE OF CONTENTS (cont.)

LIST OF FIGURES

LIST OF FIGURES (cont.)

LIST OF FIGURES (cont.)

LIST OF TABLES

LIST OF TABLES (cont.)

LIST OF TABLES (cont.)

LIST OF ACRONYMS AND ABBREVIATIONS

ACK - Positive Acknowledgment

ADCCP - Advanced Data Communications Control Protocol

ADP - Automatic Data Processing

AI - Articulation Index

AM - Amplitude Modulation

AN - Analog

ANSI - American National Standards Institute

ARPANET - Packet Switch Network developed by Advanced Research Project Agency

ARQ - Automatic Repeat Request

ASCII - American Standard Code for Information Interchange

ATC - Air Traffic Control

AT&T - American Telephone and Telegraph Company

AUTODIN - Automatic Digital Network

AUTOSEVOCOM - Automatic Secure Voice Communications Networks

AUTOVON - Automatic Voice Network

A/D - Analog to Digital

BER - Binary or Bit Error Rate

BH - Busy Hour

BISYNC - Binary Synchronous

BPS - Bits Per Second

BSL - Binary Segment Leader

CCIS - Common Channel Interswitch Signaling

CF - Communications Facility

COMSEC - Communication Security

CONC - Concentrator

CONUS - Continental United States

CRC	-	Cyclic Redundancy Check
CRT	-	Consonant Recognition Test
CVSD	-	Continuously Variable Slope Delta
DAM	-	Diagnostic Acceptability Measure
DAT	-	Diagnostic Articulation Test
DCA	-	Defense Communication Agency
DCE	-	Data Communication Equipment
DCEC	-	Defense Communications Engineering Center
DCSII	-	Second Generation Defense Communication System
DDD	-	Direct Distance Dialing
DDS	-	Dataphone Digital Service
DFT	-	Discrete Fourier Transform
DIG	-	Digital
DOD	-	Department of Defense
DRT	-	Diagnostic Rhyme Test
DSN	-	Defense Switched Network
DTE	-	Data Terminal Equipment
FACS	-	Facsimile
FCS	-	Frame Check Sequence
FDX	-	Full Duplex Transmission
FEC	-	Forward Error Correction
FM	-	Frequency Modulation
FRT	-	Fairbanks Rhyme Test
FS-1033	-	Federal Standard 1033
GOS	-	Grade of Service
GSA	-	General Services Administration
HDLC	-	High-Level Data Link Control

LIST OF ACRONYMS AND ABBREVIATIONS (cont.)

HDX	-	Half Duplex
HSC	-	Host Specific Control
HSI	-	Host Specific Interface
IBM	-	International Business Machines
ICU	-	Interface Control Unit
IF	-	Intermediate Frequency
ISO	-	International Standards Organization
IST	-	Interswitch Trunks
ITS	-	Institute for Telecommunication Sciences
I/A	-	Interactive
KB/HR	-	Kilobits Per Hour
Kb/s	-	Kilobits Per Second
LPC	-	Linear Predictive Coding
MCCU	-	Multiple Channel Control Unit
MRT	-	Modified Rhyme Test
MTBF	-	Mean Time Between Failures
MTBO	-	Mean Time Between Outages
MTSR	-	Mean Time to Service Restoral
MTTR	-	Mean Time to Repair
MU	-	Message Unit
MUX	-	Multiplexer
NACK	-	Negative Acknowledgment
NBS	-	National Bureau of Standards
NTIA	-	National Telecommunications and Information Administration
N/A	-	Not Applicable
OCR	-	Optical Character Reader
PABX	-	Private Automatic Branch Exchange

PAM	-	Pulse Amplitude Modulation
PARM	-	Paired Acceptability Rating Method
PATS	-	Perceived Acoustic Traits
PB	-	Phonetically Balanced
PBX	-	Private Branch Exchange
PCM	-	Pulse Code Modulation
pdf	-	Probability Density Function
PSN	-	Packet Switched Node
QAM	-	Quadrature Amplitude Modulation
QOS	-	Quality of Service
QUART	-	Quality Acceptance Rating Test
Q/R	-	Query and Response
RADC	-	Rome Air Development Center
R&D	-	Research and Development
SCIM	-	Speech Communications Index Meter
SDLC	-	Synchronous Data Link Control
SIP	-	Segment Interface Protocol
SNR	-	Signal-to-Noise Ratio
SVIP	-	Secure Voice Improvement Program
SW	-	Switch
S/N	-	Signal-to-Noise
TCP	-	Transmission Control Program
TPU	-	Transmission Preference Units
TRI-TAC	-	Joint Tactical Communications Program
TTY	-	Teletypewriter
VF	-	Voice Frequency

PERFORMANCE PARAMETERS FOR DIGITAL AND ANALOG SERVICE MODES

Martin Nesenbergs, William J. Hartman, and Robert F. Linfield*

As with many technologies, the evolution of telecommunication systems is shaped by two driving forces - performance and cost. There is a real need to bridge the gap between 'performance' as perceived by the user in accomplishing a mission and 'performance' as perceived by the supplier to minimize costs of implementation and operations. The interrelationships between user-oriented performance parameters, engineering-oriented parameters, and cost parameters ultimately define the permissible tradeoffs.

This report covers one phase of a multiple phase project to relate the performance needs of military network users with the performance provided by a particular telecommunications service and to seek least cost systems that would offer such service.

In this first phase, the parameters describing the performance of two services, one analog and one digital, are defined and specific values assigned for the related service offerings. The interrelationship between the user-oriented parameters and values and the technical or engineering-oriented parameters and values was planned as the subject of a subsequent phase.

Key Words: analog communications; digital communications; performance parameters; performance standards; service modes; system design; user requirements

1. OBJECTIVE AND SCOPE

Technical criteria or engineering performance specifications serve as the basis for designing various elements of a telecommunications network including the terminals, the switching nodes, and the transmission links. The cost effectiveness of the network therefore is highly dependent on how well the technical performance specifications are defined. A substantial impact on cost-effectiveness can be realized by providing more precise definitions based on user's needs, and thus preventing overdesign. This is the ultimate goal of a three-phased program of the Defense Communications Engineering Center (DCEC). A brief project overview describing the three major phases leading to a determination of digital transmission technical criteria is given in Appendix A. The specific objectives for each phase are summarized as follows:

*The authors are with the Institute for Telecommunication Sciences, National Telecommunications and Information Administration, U.S. Department of Commerce, Boulder, Colorado 80303.

Phase A - User Criteria. Develop user-oriented performance criteria for two representative services: a digital voice communication service and an interactive data communication service, for the second generation Defense Communication System (DCS II).

Phase B - Technical Criteria. Translate the user-oriented performance criteria into corresponding technical (engineering-oriented) criteria on an end-to-end basis.

Phase C - Subsystem Criteria. Develop user-oriented and technical performance criteria for additional DCS II service modes. Allocate the technical criteria to subsystems within a defined DCS II global reference network.

This report covers only Phase A, the development of user-oriented parameters and the assignment of values to them for two service modes - one analog and one digital.

In the following introductory subsections we give some background for the program, develop the basic concepts and definitions, define the objectives of Phase A, and summarize the report sections which follow.

1.1 Background

Telecommunications networks are undergoing revolutionary changes created by dramatic advances in switching, transmission, and terminal technology. The changes are driven by a demand for new and specialized services, higher quality for these services and lower cost. The new technologies are exemplified by digital carrier transmission, computer communications, packet switching, common channel signaling, microprocessing and large scale integrated circuitry. One result is that basic network elements (i.e., terminals, switches, concentrators, and transmission links) are progressively becoming more digital.

The Defense Communications Agency (DCA) has undertaken a number of actions leading to the development of advanced network architectures which take advantage of digital implementations, offer military advantages over the current systems, and reduce cost for service.

Future concepts generally envision a DCS which is secure, fully digital, and capable of handling various classes of service (e.g., voice, data, and imagery) in an integrated manner.

The forecasted evolution of military communication networks from analog, to analog-digital hybrids, and ultimately to an all-digital configuration has already begun. Digital techniques, currently being implemented on some transmission facilities at the strategic backbone level, will progress to the tactical level with the introduction of TRI-TAC equipment, and to the local military base nontactical level over the next decade. As this progression from the present architecture, to the goal architecture continues, it is anticipated that the Automatic Voice Network (AUTOVON), the Automatic Digital Network (AUTODIN), and the Automatic Secure Voice Communications Networks (AUTOSEVOCOM) will be further integrated as system improvements are made. The future Defense Switched Network (DSN) could eventually consist of a single, unified system of network elements capable of handling all of the telecommunications traffic needed by the military departments. All information signals traversing this network would be in digital form.

Since most of the digital elements of the future DCS network must still be developed and implemented, the opportunity exists to influence the technical specifications for the design. If these specifications can be defined more precisely than has been done in the past, a considerable cost savings may be realized.

1.2 Basic Concepts and Definitions

The process of transferring information over a distance is telecommunications. Many classes of telecommunications service are being offered. The type of service depends on such factors as 1) the users' demands in terms of the kind of information to be transferred, 2) the information-handling capabilities of the system supplying the service (its functional attributes), and 3) the physical structure of the system itself.

Some discussion of basic telecommunication system concepts and definitions of the general terminology used in this report is given here to aid in understanding what follows. Other, more specific definitions are given in the body of the report and in Appendix B.

In providing needed communication services to its users, a telecommunications network must meet certain functional objectives. These networks can be distinguished either by the functions performed or the services provided, as depicted in Figure 1. The services are related to the user, whereas the functions are related to the network elements, i.e., the terminals, the switching nodes, and the transmission links. The blocks in Figure 1 are defined in the following paragraphs.

The <u>end user</u> is an individual or a computer program that either produces, stores, processes, or ultimately consumes the information transferred over the system. Typical end users are the calling and called parties in a telephone conversation, the human terminal operator, the remote computer application program

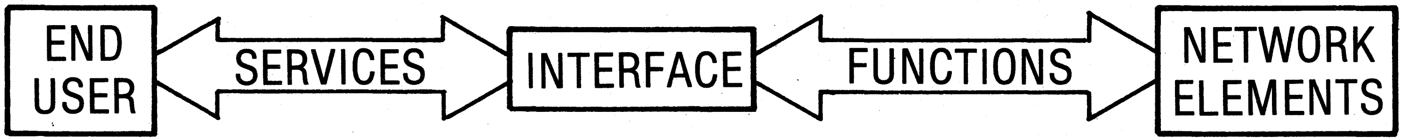


Figure 1. Basic system concepts.

in a teleprocessing network, and the electromechanical sensing, monitoring or regulating devices in a process control system.

The services are defined in terms of features which are directly observable by the end users, such as types of information sources and their characteristics. Various performance parameters may describe the services delivered to the user. Digital services are all services in which signals transmitted across the interface are represented by a finite number of descrete levels or states. Analog services include all other services. In particular, they include analog signals sampled at discrete times, but not amplitude quantized as in pulse amplitude modulation (PAM).

The interface may provide media conversion for transmitting and receiving information. Printed words or numbers, visual displays, holes in paper tape are transformed to and from electrical signals at the user interface. Information units which cross both user-system interfaces cause a discrete change in one or both communicating entities. Overhead bits on the other hand normally do not cross both interfaces. Overhead is required to manage the network functions.

Network functions define what the system must do in terms of operations and the performance of equipment or facilities. The precise definition of these functions is the network architecture or the rules upon which the network operation is based. Network architectures are often specified in terms of protocols for the pairwise interactions between pairs of similar network elements.

The network elements comprise the implementation of the functions or architecture. They describe the physical structure of the systems and subsystems used to perform the necessary functions in terms of software code, hardware components, link topology, and node switching or concentration.

The concept of an aggregate user in contrast to an end user is also useful. An aggregate user consists of an end user plus one or more adjacent system elements which collectively receive communication service from a subsystem. This distinction is depicted in Figure 2 for two service modes. In both of these modes the various circuit elements generally occur in pairs and the interfaces are identical for these pairs. For example, modems, communication security (COMSEC) units, and analog-to-digital converters all communicate with their own kind only. For end-to-end users with analog sources on the user's side, the first order and higher order parameters on the user's side are different from the technical parameters on the system side. With digital sources on the users side the first order parameters may be the same.

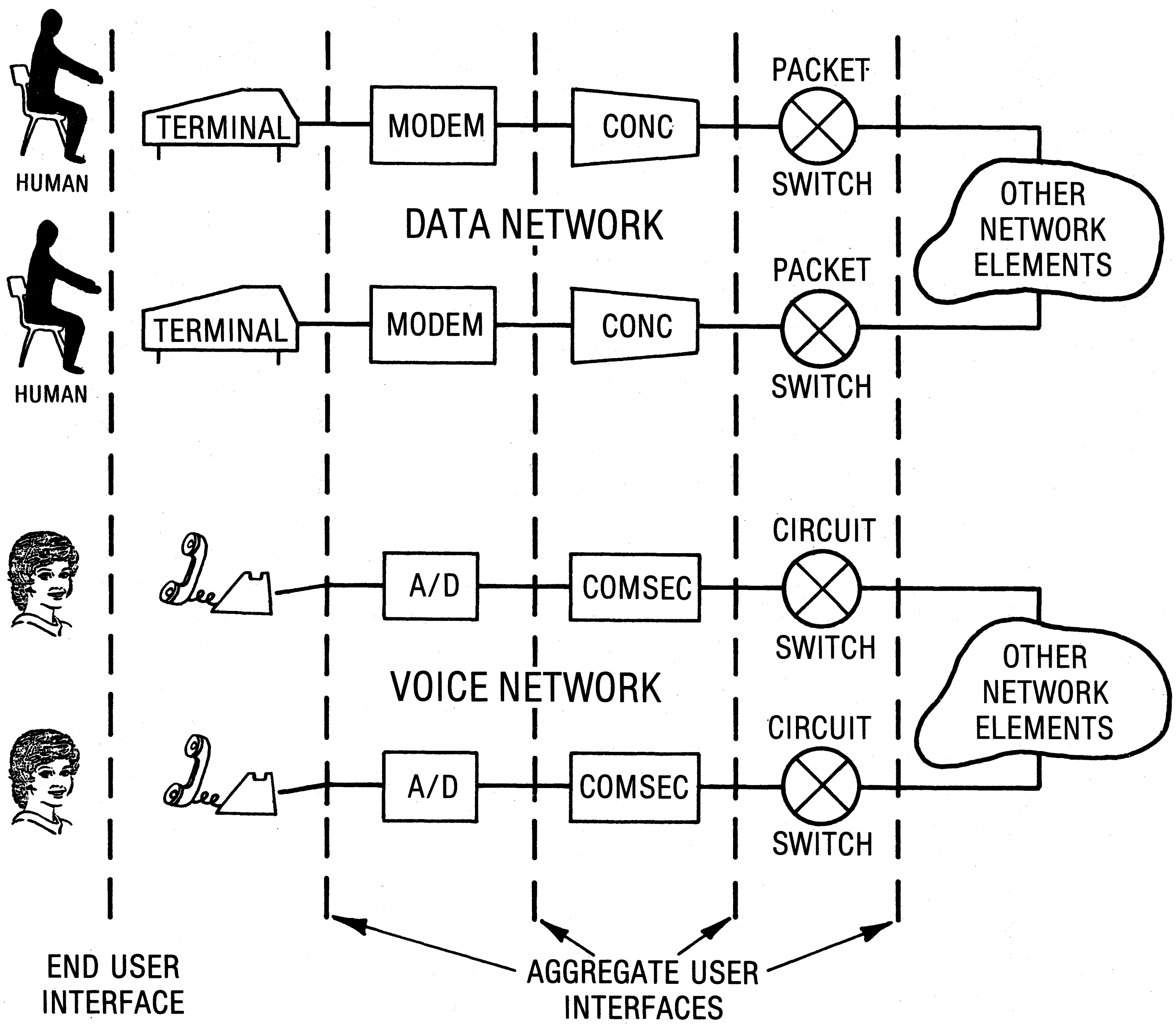


Figure 2. Distinction between end user and aggregate user interfaces.

There are three phases to the process of telecommunications. The first, called the access phase, consists of a number of activites for provision of access paths so that the end users can eventually transfer information. Once the access path has been provided the information transfer phase begins. This phase may be unidirectional or bidirectional, so that a monolog or dialog between end users can be carried out. Completion of the transfer phase marks the beginning of the disengagement phase, which ends when the system is returned to its initial state and is ready to allocate the access path to other users.

Access paths may be continuously available (using dedicated circuits) or they may be set-up and terminated on an intermittent basis (using circuit switches or store-and-forward switches). The patterns of intermittent information exchange can vary from a few groups of information bits, such as characters, to complete words, blocks, frames, packets or messages. They may be transmitted in just one direction or in both directions. The number and frequency of transactions can vary. Table 1 lists some examples of end-user interactions for various network configurations.

It is apparent from Table 1 that many different parameters and values are required to characterize system performance of different systems for all three phases. Some are needed to depict successful performance, others for incorrect performance, and finally for nonperformance. In order to cover all three categories both long- and short-term measures of performance may also be required.

1.3 Objectives

Like many technologies, telecommunications networks are shaped by two driving forces: performance and cost. A given network, more often than not, is a compromise between these two conflicting forces. For instance, a common carrier network user desires efficiency, accuracy, and reliability. In particular, he may desire fast access with service always available. This user also wants to minimize costs by using the network as little as is necessary to meet his needs. The common carrier network supplier, in contrast to the user, has economic reasons to sell as much service as possible; to produce more income by sharing the network resources among more revenue generating customers.

The military user of a network has many of the same needs as the civilian user, but requires a higher degree of reliability and availability. In addition, the military user has certain essential requirements such as mobility, security, and survivability. The military user's service needs are different and the network he uses must be adaptable to more dynamic conditions. The military network

Table 1. Examples of End User Interactions

Form of Dialogue	Packet Switched Network	Circuit Switched Network	Message Switch Network	Television Broadcast
Datagram	Single Packet Unidirectional	Single Word	Single Block	Single Line
Transaction	Packet Elicits Response of Fixed Number of Packets	Sentence and its Response	Several Blocks	Complete Frame
Session	Bidirectional Packet Flow Series of Transactions	Complete Call	Message Delivery	Sequence of Frames

operator may also be viewed as a paying consumer rather than a producer (collector) of revenue. Such operators must provide adequate service in a most economic fashion. This economy can only be realized by careful consideration of all the network elements, including cost to make, own, operate, and maintain the terminals, the switching nodes and the transmission links. The importance of minimizing costs cannot be over-emphasized. Almost one billion dollars per year is expended on new installations, and on the operation and maintenance, for the Defense Communication System (DCS) alone. This is more than one fifth of the yearly total expenditures for communications by the Department of Defense.

A principal objective of this project is to relate the performance needs of a military network user to the performance provided by a particular service, and to seek least cost systems that would offer such service. The military departments are the users of concern. The system supplier is the second generation Defense Communication System (DCS II). Military users require certain capabilities pertaining to the effectiveness of information transfer for voice, data, and imagery. Although given classes of users know telecommunications applications well, they seldom understand the technical aspects of communications. Conversely, the network designer or supplier understands the underlying technologies, but has incomplete knowledge about a particular user's application. If cast in the role of network designer, the user tends to generate overdesigned, inefficient, and costly networks. Suppliers also tend to overdesign because they are uncertain about the user's real needs. The upper portion of Figure 3 depicts how the communications user usually specifies his service needs in general terms based on mission requirements. The network designer then develops functional requirements for the network and prepares engineering specifications for the various network elements, i.e., the terminals, the links, and the nodes.

User panels can evaluate the system using subjective performance measurements to see if their requirements have been achieved. Performance evaluations by the network designer are also helpful to verify whether the engineering specifications are met. All too often this procedure results in either overdesign or underdesign on the part of the supplier. Thus, there is a real need to bridge the gap between users and suppliers by establishing a relationship between user-oriented parameters and engineering-oriented parameters as shown on the lower part of Figure 3. This matching process is required for different classes of users, for different services and for various modes of operations. More cost-effective systems are the result.

The parameters describing the performance of services delivered to an end user are the subject of this Phase A report. Values must be assigned to these

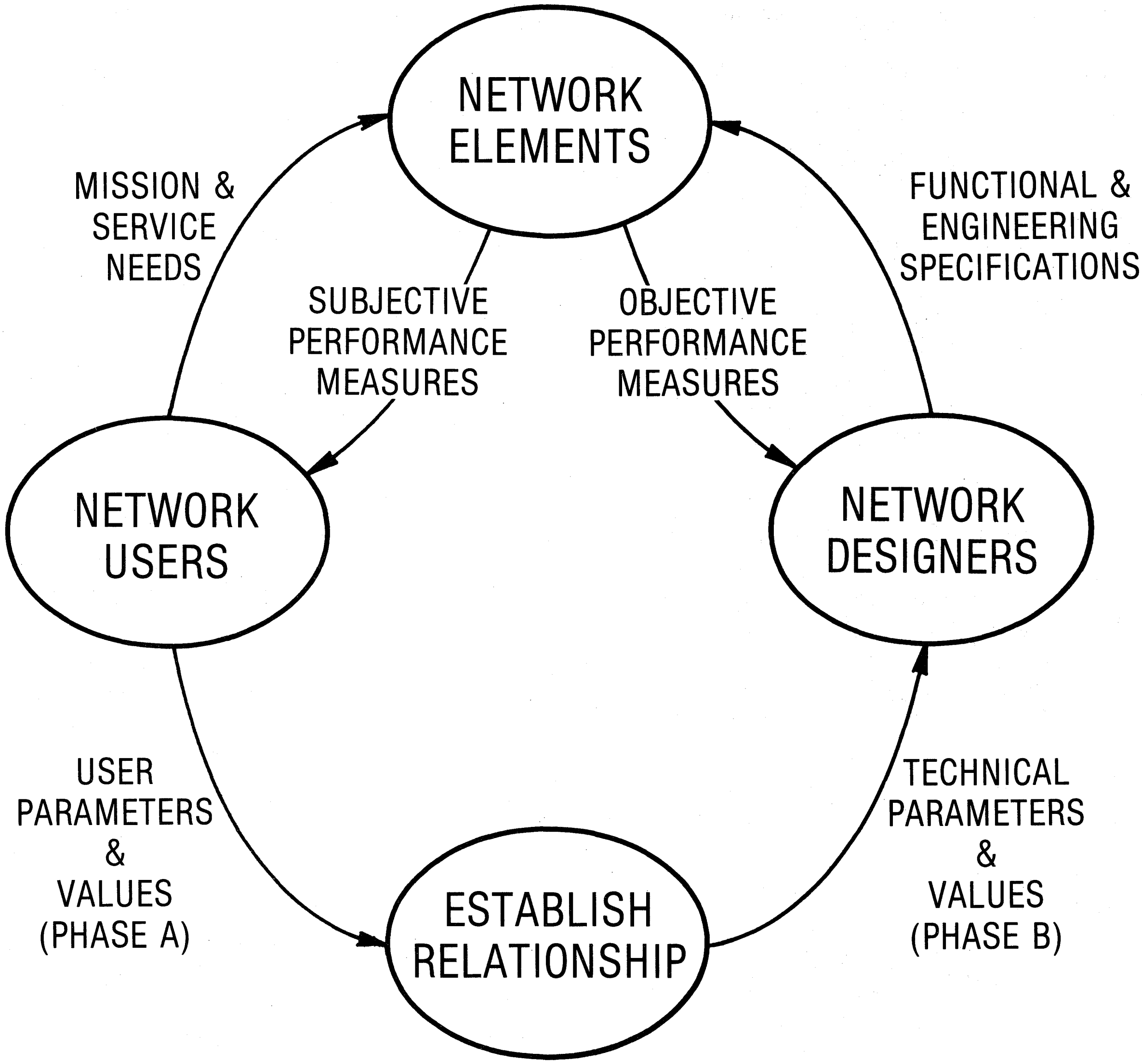


Figure 3. Project objectives.

user performance parameters for selected service offerings. Measurements of parameters on operating systems can be used to indicate when the performance is satisfactory and when it is degrading.

The interrelationship between the user parameters and values, and the technical or engineering-oriented parameters values must be established later. Technical performance parameters are essential to the design, development, construction, operation, and maintenance of the entire system. Values assigned to the technical parameters of the system elements are important to both internal operation and service. Measures of these parameters can be used to indicate where and why the performance is degrading.

In order to achieve the objectives of Phase A a sequence of steps is required. Each step is a prerequisite to the subsequent step. The steps are as follows:

1. Select two communication services, one analog and one digital, from representative service types available on DCS II networks.

2. For each service selected define parameters which describe the performance from the end user's viewpoint.

3. Identify DCS II operating modes for each communications service in terms of transaction profiles.

4. Chose appropriate subsets of user-oriented parameters for each service mode.

5. Define mission categories to distinguish classes of users having the same or similar performance requirements.

6. Assign numerical values to each parameter subset based on a specific user class.

It is apparent from these steps that the scope of Phase A should be limited in order to accomplish Phase A within a reasonable time frame. The present study is limited to only two types of service and one mode of operation for each type. The emphasized parameters are those that are expected to have the greatest impact on transmission facility design. Finally, the numerical values will be assigned for only one user class.

We are still confronted with a sizable task. Fortunately, a great deal of work has already been done to develop and define user-oriented parameters for digital services. See for example, Seitz and McManamon (1978). One result of this work is the interim Federal Standard entitled, "Digital Communications Performance Parameters," and designated FS-1033. A closely related standard entitled, "User-Oriented Data Communication Performance Parameters," is currently being developed by the American National Standards Institute (ANSI) and is designated X3S35/125. These prior efforts provide the essential data base needed to expedite this phase. Because of its importance to this study, a published description of FS-1033 is reproduced in Appendix B (Seitz and Bodson, 1980).

Both the interim standard FS-1033 and the proposed standard ANSI X3835/125 provide a user-oriented means of specifying performance in a system-independent way. The standards achieve this by placing the user-system interface between the end user and a data terminal. However, these standards may also be used to specify the performance of facilities or services which are terminated within the end user interface (e.g., the aggregate user concept), providing that these interfaces can be described in digital terms. This is an important point, since internal digital interfaces are often found in networks supplying analog as well as digital services. The interface following analog-to-digital conversion in a voice circuit is one example.

The standards do not identify numerical values for any of the performance parameters. The values may be specified by the user to indicate the minimal requirements of the desired service. Or some nominal values may be specified that exceed the minimum requirements and do not impair the systems practicality.

The methods for measuring data communication system performance parameters of FS-1033 are not included in the existing standard. That is the subject of a separate standard currently under development.

The analog service parameters selected follow generally, if not identically, the same path as indicated for digital service in FS-1033. A session between users in this case may be a telephone call. A 'call' is divided into the three functional phases of access, information transfer, and disengagement. Each phase can result in one of three performance categories - successful, incorrect, and nonperformance. Thus, there are again nine separate groups of parameters needed. More than one parameter may be required to characterize the performance in each group. The access and disengagement phases can apparently use the same set of parameters as used in FS-1033. This leaves three information transfer performance categories to be defined for the analog service.

Figure 4 summarizes the approach used. User-oriented parameters are selected first, followed by selection of their numerical values. One of the more difficult tasks in Phase A is to define the end-user parameters for the analog service that employs digital transmission of voice. Although a vast amount of work has been done in this area (see, for example, Flanagan et al., 1979) it is difficult to specify

		USER-ORIENTED PERFORMANCE PARAMETERS			PARAMETER VALUES FOR SELECTED MODES OF OPERATION		
		SPEED	ACCURACY/ QUALITY	RELIABILITY	SPEED	ACCURACY/ QUALITY	RELIABILITY
DIGITAL SERVICE	ACCESS						
	TRANSFER		SUBSETS CHOSEN FROM FS 1033 PARAMETERS			TO BE ASSIGNED FOR SELECTED AUTODIN II MODE	
	DISENGAGEMENT						
ANALOG SERVICE	ACCESS		SUBSETS CHOSEN FROM FS 1033 PARAMETERS				
	TRANSFER		TO BE DETERMINED			TO BE ASSIGNED FOR SELECTED AUTOVON II MODE	
	DISENGAGEMENT		SUBSETS CHOSEN FROM FS 1033 PARAMETERS				

Figure 4. Phase A approach.

voice quality quantitatively. This is so because the quality of voice signals implies measures of fidelity. Fidelity measures involve human listener perception of what is said and who said it. Many different subjective methods for quantifying voice quality have been tried using various types of listener panels. Objective measures of perception are currently actively pursued because they are easier to implement, and at less cost. It is difficult, however, to obtain general acceptance of objectively measured scores by voice system designers. These aspects are discussed in more detail in Sections 6 and 7.

1.4 Parameter Selection Concepts

This report deals with the concepts of system performance, the selection of technical parameters to represent the system performance, and the assignment of values for the parameters which are eventually selected. This consideration of system performance applies at the interface between the user and the system.

Since system designers are not frequently required to apply system requirements to the user-system interface some emphasis must be given to its location. In this report, the user-system interface is considered to be the end of the telecommunication function. The location of the user-system interface must be kept in mind while reading this report.

At the user-system interface, one has two points of view which must be reconciled. The user's viewpoint can be expressed by looking toward the interface from the user's side and observing the received service. At the same time, the system designer's viewpoint is obtained by looking towards the same interface from the systems side and observing the offered service. For interface compatibility to exist relative to performance definition and assessment, the user's set of performance parameters must be contained within the system designer's set of performance parameters. Conversely, the system designer's set of performance parameters is not necessarily limited to the user's set.

As clarification of this point, let the set of user performance parameters on the user side of the user-system interface be designated as set A. Then the system designer's set of performance parameters on the system side of the user-system interface would consist of set (A + B).

Set B consists of those parameters added by the system designer at the user-system interface which the designer considers necessary in order to satisfy set A or to operate the system compatibly with set A, or both. Set B need not be known to the user. Most often, set B is not known to the user. On the other hand, in some applications, the system designer does not add any parameters to the set A and the set B is the null set. In summary, the user parameters (set A) are necessary but not always sufficient to describe performance at the user-system interface. Since it depends on the system, set B is not used here.

In this study, interim Federal Standard 1033 was reviewed for application here. The performance parameters of FS-1033 have been defined therein as a minimum set of parameters applicable at the user-system interface from the user's viewpoint. In FS-1033, average values are specified as measures of the parameters. The study incorporates the FS-1033 average value parameters as part of the previously mentioned user performance parameter set A. The importance of average values extend to both sides of the user-system interface in this application study.

One must recognize that all the parameters considered for selection are intended to measure in some way the behavior of a multidimensional random process. The parameters are random variables defined on sample functions of the random processes. Complete specification of performance, once the parameter is selected, does require attention to the issues of statistical stationarity and ergodicity as well as definitions of the random variable distributions. Once distributions are properly defined, means, averages, variances, and higher order moments can be sought.

Consider for illustration the FS-1033 parameter access time. For convenience in this discussion, let t_a designate access time. Based upon the definition of the parameter 'access time' and the interface definitions, a probability distribution of t_a is defined, represented as $P_a(t_a \leq T_a)$. Similarly, one has available, by definition, the average access time $E(t_a)$, the standard deviation $\sigma(t_a)$, and various percentile levels, such as $PL = P_a^{-1}(0.9)$ for the 0.9 percentile. Obviously, other parameters can be defined.

More generally, one may be faced with time variations which must be recognized, such as busy and nonbusy hour intervals. If one designates these intervals of time as I_j, the jth interval has $P_a(t_a \leq T_a \mid I_j)$, and all the other parameters are conditional upon I_j as, for example, with $E(t_a \mid I_j)$.

This aspect of parameter selection is important because the parameters must be definable in a manner that allows measurement as quantitative variables. The ability to measure is very sensitive to the interface chosen.

At the user-system interface as defined by FS-1033, the observables available to the user must be identified carefully so that they can be measured or, at least, calculated unambiguously from other measured values. A short example is helpful here for illustration.

The parameter 'connection time' is more conventionally used than access time. To measure connection time, however, requires detection of the last digit of a dialed number as it leaves the originating caller's station apparatus (to start the measurement of connection time). The end of connection time is given by the detection of the initial audible ringing at the originating caller's station. Although both of these events are measurable, their measurement by users, and particularly application programs in computers, is quite difficult because these events are not easily observable to the user.

The start of the access time interval, however, is easily measurable by the user. The end of access time is denoted by the transfer of the first information bit for the session across the user-system interface. This end time event is not always easy to detect, but it is more readily observable and programmed, particularly in application programs and terminals.

The user-system interface, as with all interfaces, has the function of identifying a point at which commonality can be defined for use and reference by both sides of the interface. However, this is a minimum condition when parameters which are random variables are involved at the interface, as noted before relative to the FS-1033 parameters.

The parameter access time is specified in interim FS-1033 as an average value on the user side of the interface, namely, $E(t_a)$. Thus, on the system side of the interface, the system designer and operator must be concerned with $E(t_a)$. However, as far as interim FS-1033 is concerned, $\sigma(t_a)$ could also be specified on both sides of the interface if it is considered of value to the users, or just on the system side if the system designer considers it important. Conversely, however, one cannot specify $\sigma(t_a)$ just on the user side without including it on the system side.

Going farther, access time can be subdivided on the system side of the interface into its constituent component parameters which are dependent upon the particular mode and application. For a switched network, conventionally dialed telephone call, these component parameters can be defined as, for example,

- o user dial time interval (calling user)
- o connection time interval
- o user answer time interval (called user).

Of these component parameters, the connection time is certainly of interest and value to the system designer. The effect of the other two are also of considerable interest and value to the user.

The statistical aspect of the parameter selection process is considered in more detail in subsequent sections. The special requirements of users in this application have led to the conclusion that the performance parameters on the users' side of the interface should not be limited to the average values, but should reflect some measure of both dispersion about the mean and a worst case or failure likelihood associated with performance. For these reasons, the 90 percentile has been used whenever applicable. Hence, the previously cited parameter access time is specified as having two values of interest, the average and the 90 percentile.

A final concept associated with parameter selection is the specification of values for the various average and 90 percentile parameters. In the larger sense, values of performance parameters most often represent a compromise between user requirements based upon mission and functional requirements, costs, and the practical limitations of current systems either in operation or in development. Obtaining user requirements often requires an iterative process which begins in the Concept Development Phase and is partially resolved in the following System Validation Phase.

This iterative process is not within the scope of this study. Nor is it appropriate to analyze costs or mission requirements, however noteworthy such analyses may be. It was planned to use existing measured data on system performance. Unfortunately, as the reader will note, such data are not abundantly available. As the report describes, all the numbers chosen for parameter values were chosen carefully. The lack of data, and validation tests, strongly suggests that the numbers specified are to be interpreted as guidelines to illustrate the relationship between parameters and to form the basis for validation.

1.5 Report Synopsis

Section 2 reviews methods for classifying communications services from the viewpoint of the traffic manager, system designer and the end user. Various service classification schemes have been used by these three groups. For instance, the traffic managers have in the past emphasized traffic classes, volumes, and congestion statistics. The system designers, by necessity, have taken a technical or engineering approach to all elements of transmission links, switching nodes, terminals, and the architectural structure that ties them together. The objective here is to stress those service aspects that are perceived by the end users. It is important initially to categorize and divide both services and users into groups having similar functional, performance and application characteristics. This will ensure, as one expects, that the same parameter set will apply to the users in question.

Different types of telecommunications traffic have different military significance, as well as different volumes of flow. These characteristics are further categorized for selection of operational modes to be studied here. This selection of modes also includes the main functional aspects of the three principal network elements (i.e., the links, nodes, and terminals). Tabulation of their network properties will be needed eventually.

The selected digital service mode, to be studied subsequently, involves a typical host computer application program communication with either another host or with a sophisticated high speed terminal. The latter may have a human operator interaction. The analog service mode, to be selected, involves voice or audio sessions between humans in a secure conferencing (e.g., more than two active terminals) arrangement. Section 3 discusses the procedure used for the selection of operating modes for these two service classes. Two basic principles have been instrumental to the mode selection process: the service utilization level and the performance vulnerability to backbone transmission impairments. The latter parametric criterion appears particularly relevant, since technical performance levels developed for the most critical services should provide satisfactory service for other less critical modes.

The following digital service mode of operation has been selected from the various AUTODIN II interactive (I/A) modes. It is a high speed, 56 Kb/s, synchronous, binary, full-duplex mode with the Advanced Data Communications Control Protocol (ADCCP). The transmissions employ data packets or segments. The segments are exchanged rapidly during specially setup virtual circuit sessions.

The analog service mode selected is a digital voice mode. It uses a Continuously Variable Slope Delta (CVSD) waveform encoder, plus a 16 Kb/s modem. The mode permits conference calls to be secure along the lines of the Secure Voice Improvement Program (SVIP).

The characterization of the two selected modes starts with simplified versions of their transaction profiles. As the project proceeds, more detailed profiles may be needed to facilitate the mapping of user performance values into system performance values. Several key subsystems may have to be defined to realize this mapping. During the initial user-oriented phase of this program, however, detailed attention to these technical values is not needed. What is required is an assessment of realistically available services with existing technology.

Many service parameters have been used in the past and many are used today. Section 4 begins by scrutinizing some 66 typical parameters. With different relative weights, these parameters depict the different data communications performance aspects of service, systems, and network operation. Since we are primarily interested in user-oriented performance parameters, and fewer in number, Section 4 seeks to reduce the parameter set to something adequate for the military service users. The best available and most researched digital parameter set appears to be the Interim Federal Standard FS-1033 (or FS-1033, for short). A description of FS-1033 is attached here as Appendix B. FS-1033 defines 26 digital communication performance parameters. This total set may be compared with any other set of popular user-oriented performance descriptors. Section 4 presents such a comparison with 16 rather common user-oriented performance parameters that are particularly applicable to the AUTODIN II digital service modes. In several respects, FS-1033 is found to be equal or superior to the 16 popular parameter set.

Using the definitions of performance parameters, Section 5 proposes quantitative parameter values for the selected digital service mode. Considerable discussion appears warranted here. First, the selected mode must be adequately characterized. Then numerical values must be developed and justified for the FS-1033 parameters. As noted in the respective sections of this report, the communication transaction is divided into three phases: the access phase, the information transfer phase, and the disengagement phase. Each phase is represented by several FS-1033 parameters. For instance, both end users and system engineers appear concerned with marginal or unacceptably deteriorated performance as well as with the typically acceptable average performance. Viewed according to orders of magnitude, the unacceptable performance may occur less than 1% of the time, whereas the acceptable, "near average", performance may be observed most of the time. Both average and 0.9 percentile values are recommended to be used for the parameters.

Because of the scarcity of empirical user-oriented service requirement data, our suggested numerical values are entirely too speculative. The reader should view them as initial benchmarks to be critiqued, modified, and improved. The values listed in this report, however, do fit the meager experimental data base and are reasonably consistent with each other.

Section 6 deals with the identification of service parameters for the selected voice service mode. Emphasis is placed on the information transfer phase, because the access and disengagement phases appear moderately close to the previously discussed phases of FS-1033. By necessity, the characterization of voice qualities during the transfer phase calls for a different approach. Although the technical literature is replete with innumerable means of evaluating voice services, no single method has acquired any general acceptance. Section 6 starts with an overview of the existing performance parameters for voice systems and services. Many definitions and measurement techniques are found, but no apparent standardization.

The lack of generally accepted parameter definitions seems intimately related to the fact that subjective, simple, reliable, and repeatable means have been nearly impossible to produce, despite a great R&D effort dedicated to this task. The ongoing changes in voice test procedures and experiments, with both listener panels or opinion polls, indicate a lack of stability and perhaps progress in the subjective testing area.

Given the above state of affairs, the initial study effort reported here seeks objective measures that satisfy the five criteria: reliability, repeatability, usability, system independence, and user orientation. The principal voice channel qualities sought in Section 6 are (1) intelligibility, (2) speaker recognition, and (3) broad acceptance by users. To acheive these goals, Section 6 introduces and discusses five objective voice quality parameters. They are:

1. A normalized energy measure;
2. A short-term signal-to-noise ratio measure;
3. A bandweighted signal-to-noise ratio measure;
4. A log area ratio measure;
5. A speaker recognition measure.

The current status of subjective interpretation (intelligibility, acceptability, recognizability) to these five measures is also indicated in Section 6. Results indicate that additional data are needed to quantify the objective parameters in terms of users needs.

Where feasible some values are given in Section 7 for the access and transfer phase of a voice service using the 16 Kb/s CVSD mode of operation. The table summarizing available values for objective measures and subjective interpretations of this mode again illustrates that a great deal of additional data is needed to specify values for appropriate user groups. In some cases even ranges of values for a given parameter/system set are not available and much work remains to be done in this area.

In Section 8 we describe some methods for obtaining a data base of user requirements and means to select values for specific user groups.

1.6 Summary of Results

The key results of this study consist of user-oriented performance parameter identification and value assignment for two selected service modes. The selected digital service mode (see Section 3.2) is an interactive, 56 Kb/s, Mode VI service with ADCCP protocol. The analog service mode (see Section 3.3) is 16 Kb/s CVSD digitized voice with potential encryption and conferencing features.

The parameters recommended for the digital mode are those of Interim Federal Standard 1033 (see Section 4.2). The most significant of the 26 parameters defined in FS-1033 are the 19 so-called primary parameters. Their names and recommended numerical values for the above selected digital mode are as follows:

(1) Access Time: 0.10 s (Mean)
0.15 s (0.9 Percentile).

(2) Incorrect Access Probability: 10^{-10}.

(3) Access Denial Probability: 10^{-3} (at 0.3 s).

(4) Bit Transfer Time: 0.5 s.

(5) Bit Error Probability: 10^{-10}.

(6) Bit Misdelivery Probability: 10^{-11}.

(7) Extra Bit Probability: 10^{-11}.

(8) Bit Loss Probability: 10^{-11}.

(9) Block Transfer Time: 0.5 s.

(10) Block Error Probability: 10^{-9}.

(11) Block Misdelivery Probability: 10^{-9}.

(12) Extra Block Probability: 10^{-10}.

(13) Block Loss Probability: $3(10^{-11})$.

(14) Bit Transfer Rate: 8510 b/s.

(15) Block Transfer Rate: 8510/n blocks/s, where n denotes the number of bits per block.

(16) Bit Rate Efficiency: 50%.

(17) Block Rate Efficiency: 50%.

(18) Disengagement Time: 0.05 s (Mean)
0.10 s (0.9 Percentile)

(19) Disengagement Denial Probability: 10^{-3} (at 0.15 s).

Detailed definitions, discussions, and numerical summaries of these digital parameters are found in Section 5.

In general, the performance parameters for the analog voice mode are different from those selected for the digital mode. However, there are both certain noteworthy similarities and differences. For example, the access and disengagement phases of both the analog and digital modes can employ the same, or nearly the same, set of user-oriented service parameters. This is in sharp contrast to the information transfer phase, where the analog mode parameters bear no resemblance to those of the digital mode. In this study, starting from a large initial set of analog performance parameters, four final transfer phase parameters are selected. They are intelligibility, acceptability, recognizability, and round trip delay. The complete set of performance parameters selected for the voice service mode are listed in Table 50.

Table 50 also indicates the availability of actual values for these parameters when applied to two particular system types. Table 51 indicates the availability of corresponding values for military and other user applications. These tables reveal major gaps in the existing measurement literature in both analog performance categories. A few actual values are given in Table 52 for two existing telephone networks, the European AUTOVON network and the direct distance dialing (DDD) network.

Values given are complete for the digital mode and incomplete for the analog mode. For both modes, however, these values are predominantly "system-oriented" in that they reflect primarily levels achieved with existing technology rather than what users actually need. This absence of purely "user-oriented" values is a fundamental limitation to the work reported here.

While the development of purely user-oriented parameter values was not possible in this study, that objective is both achievable and important. Two approaches for developing a firm "user requirements" data base are suggested in the final section of the report. One approach involves analyzing the functions performed by a user group in carrying out its mission. The other approach involves correlating objective performance measures on systems with subjective assessments of that performance. Both approaches should yield ranges of user-oriented values for the performance parameters if a suitable methodology can be developed. It appears certain that specification of minimal, not to be violated, performance values for systems will require an iterative cost/benefit analysis involving system planners, designers, users, and operators.

2. CLASSIFICATION VIEWPOINTS

There are many ways to categorize telecommunication networks. For example, they could be classified by application, by architecture, by ownership, or by the features offered, and so forth. Here we are concerned with three classification perspectives: the users, the traffic engineers, and the system designers. See Figure 5. The user is interested in the performance of services delivered rather than the performance of equipments and facilities which supply these services. Performance measures from the user's viewpoint depend on mission requirements. Thus, there is a need to classify user services and missions.

The traffic engineer is concerned with the types of terminals, their geographic distribution, and the volume and kind of traffic they generate. Based on this information various network topologies and control procedures are considered by traffic engineers to insure that the desired quality of service is achieved. The network is specified through node and link functions that meet the network design objectives. The system designer can then provide equipment and facility specifications in quantitative form. He defines the required functions in terms of engineering parameters and then numerical values.

In the following subsections network classifications are developed from the viewpoints of the user, the traffic engineer, and the system designer. Then we can select specific services and modes of operation for the performance parameter selection process.

2.1 The User

It is possible to classify user services in various ways. Stavroulakis (1972), for example, specifies a service in terms of the attributes of the switching, transmission, and terminal parts of the network. Others classify them in terms of the applications, e.g., fixed, mobile, broadcast, data, etc. The basis for defining service classes here is different. The objective is to group services so that each group has similar performance characteristics as perceived by the end user. This approach simplifies the development of user parameters and the assignment of values for specified services.

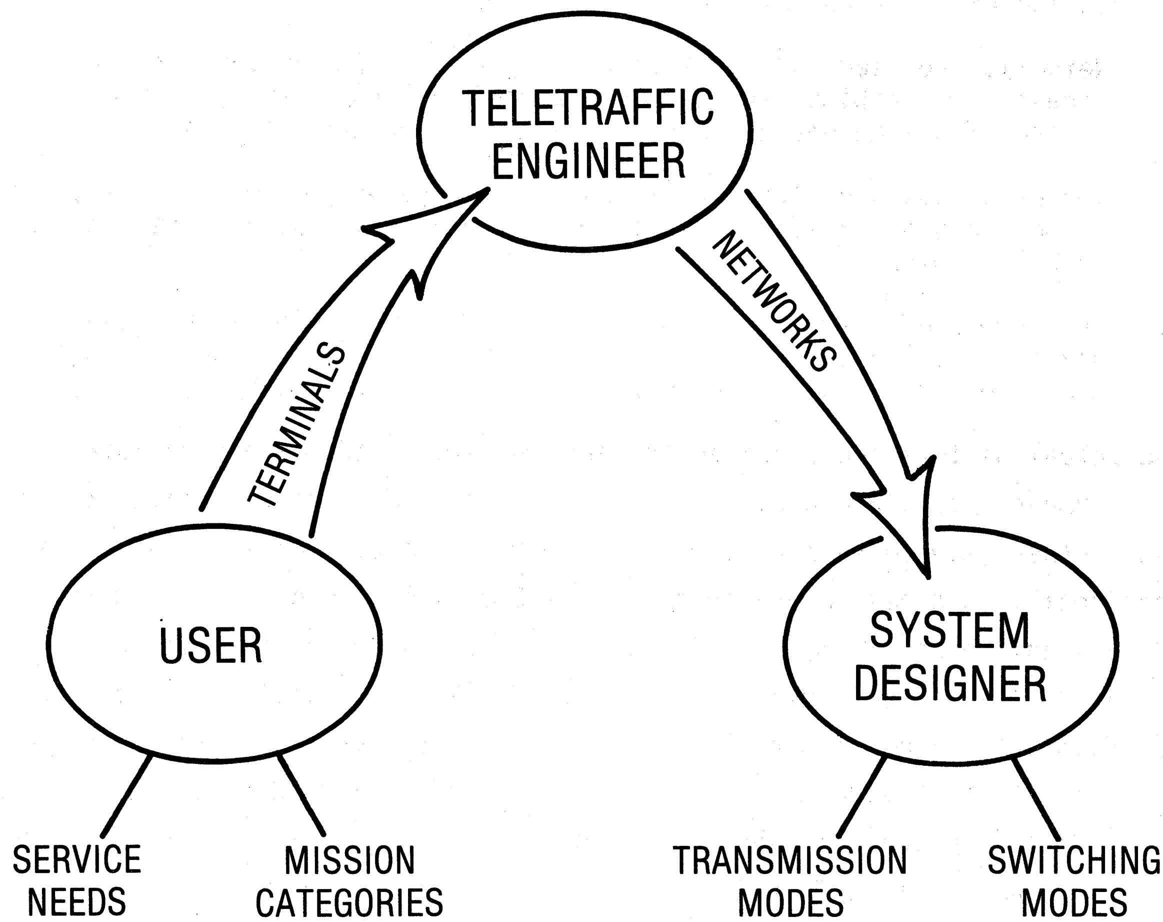


Figure 5. Classification viewpoints.

The service classifications are illustrated in Figure 6. Five major levels of division are shown. The levels are, beginning at the top of Figure 6:

1. The nature of the information signal perceived by the end user. At this level, the signals are either continuous (analog) or discrete (digital).

2. The type of source or human/machine usage of the information. For analog services, there may be audible, visual, or other sensory sources. For digital services, the sources may be a human operator, device media or computer applications program.

3. Networks are used for three general types of interaction: human access to a machine (such as a computer) and vice versa, machine interaction with one another, and interaction among humans.

4. The directivity of the access path. The information may be transferred in one direction only (simplex) or in both directions (duplex or half duplex).

5. The number of users, human or machine, that participate in a given dialogue can vary. This involves at least two or more end users on a one-to-one, one-to-many, many-to-one, or many-to-many basis.

The actual performance required for each service class depends on other factors. Some networks are designed to serve users from a single community of interest. Others may serve many communities of interest. The single-user networks are functionally specialized and optimized to the user's needs. The common-user networks are not specialized. They must be adaptive to many different user's needs. In some cases, however, the user's view of performance is highly dependent on what the user does, or the mission he or she performs. Military users of telecommunication networks can be divided into the three basic mission categories: strategic, tactical, and nontactical, as shown in Figure 7. In each category various types of information may be transferred. We have divided this information into four basic types: intelligence, command and control, operations, and administrative.

Users of military networks must often contend with the limited resources available. When contention occurs, priorities are established using multiple precedence levels. In AUTODIN I, for example, the precedence levels used in the continental United States (CONUS) are as follows:

Level	Precedence	Designator	Handling Time
I	Flash	Z	10 minutes
II	Immediate	O	30 minutes
III	Priority	P	3 hours
IV	Routine	R	6 hours

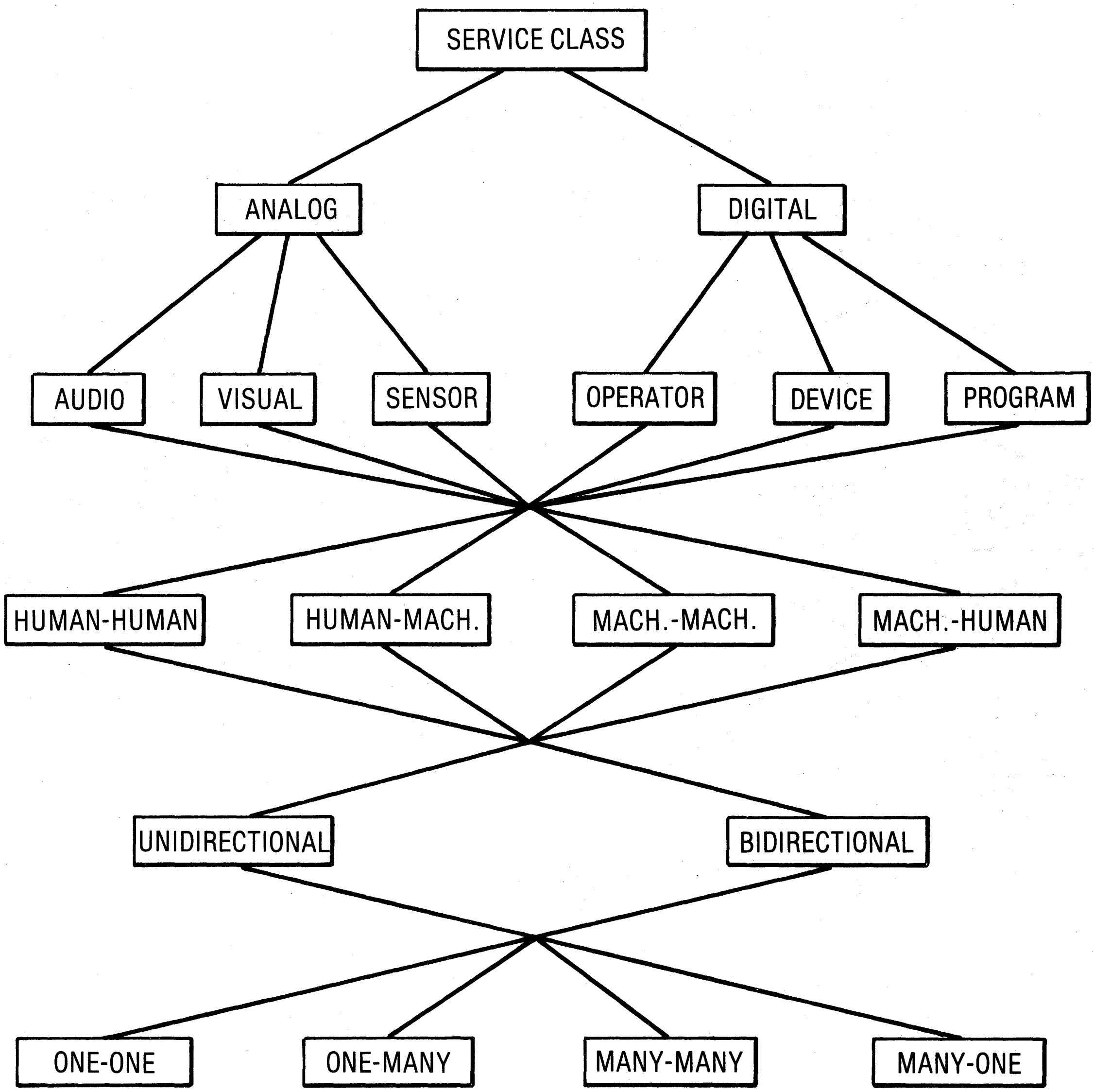


Figure 6. Service classification scheme.

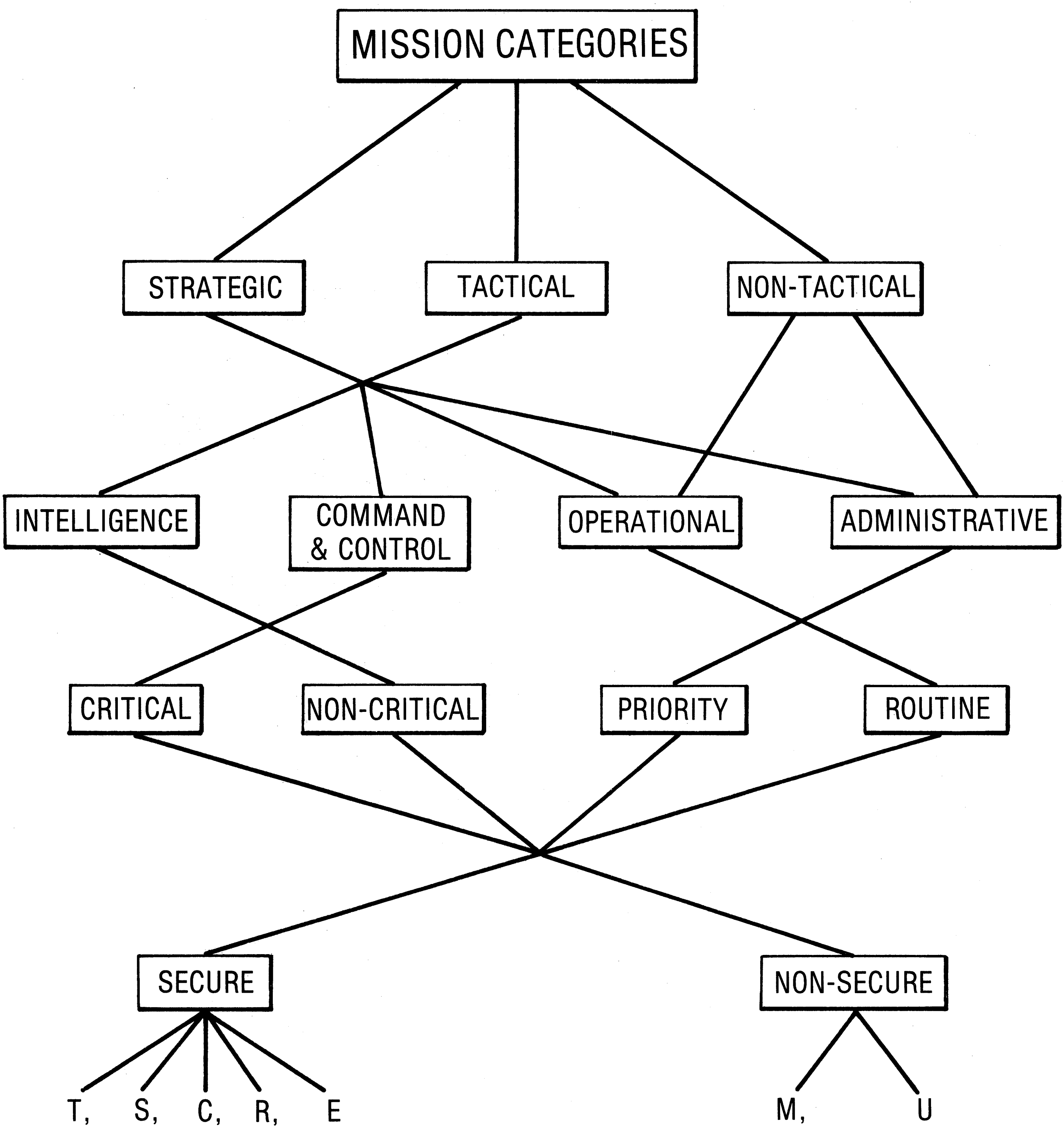


Figure 7. Mission categories.

The handling time given above is the speed of service from receipt at the origination node until delivery at the destination node. It has been estimated that 50% of the messages handled by AUTODIN I are routine messages and approximately 1% are flash messages. A 16-level precedence scheme is planned for AUTODIN II (Shah, 1977). AUTOVON I employs a flash override precedence level in addition to the four levels used in AUTODIN I. When lines or trunks are busy the AUTOVON switch can also preempt lower precedence calls in progress. Military networks also use multilevel security to prevent disclosure of classified information. These levels of security are tabulated below along with the designator used in Figure 7.

Level	Designation
Top Secret	T
Secret	S
Confidential	C
Restricted	R
Encrypted for Transmission Only	E
Classified (clear transmission)	M
Unclassified	U

The letter designators are used in message headers as part of the overhead control signals to identify the security level. When the designator is received at a node it is checked against the security classification of the destination terminal. A secure access path must be found to that terminal.

2.2 The Teletraffic Engineer

The traffic engineer is concerned with user terminals, the traffic they generate, and the means for providing acceptable access paths between them.

The type of terminals, their geographic distributions, the volume and kind of traffic generated by each terminal, as well as the required grade of service, are important factors in any network design.

There is no foreseeable limit to the types and variations of terminals. New types are continually being introduced. They range from the traditional voice terminals consisting of a telephone handset with rotary or pushbutton dial; to keyboard and printer terminals for teletypewriter and computer access; to visual display terminals with cathode ray tube or other optical readout. Many terminals incorporate microprocessors and memory. Software programs add useful intelligence and processing power to such "smart" terminals. A program addition can change the character of the terminal to meet changing communications requirements or to adapt to new applications.

The telephone handset is the most common terminal in use today. There are over 300,000 handsets in use by the military now and the number is not expected to decline. Over 500,000 telephone terminals with about 1 million telephone instruments (including extension telephones) are projected to be in worldwide use by all U.S. military services in the mid 1980's (Wagner, 1977). Data terminals are experiencing their own relatively rapid growth and could, over the same period, approach 50,000 terminals worldwide. Note that this still is only 10% of the telephone terminal population.

There are approximately 1500 military access areas worldwide. Based on the number of user terminals in each area, the areas range from small (<300 terminals), to medium (300 to 3000 terminals), to large (>3000 terminals). Although the majority of terminal types are telephones, they also include computer, teletypewriters, facsimile, and a myriad of other terminals. Terminal densities vary from less than 10 per km^2 to over 10,000 per km^2. The higher density cases offer a variety of line concentration alternatives.

The estimated number of terminals of different types for the continental United States (CONUS) and overseas, as shown in Table 2 for 1978. The average duration of the different messages generated by these terminals ranges from several hours, for bulk data transfers, to less than a minute, for voice calls, and to a few seconds, for interactive data transfers.

Table 2. Estimated Number of Military Terminals for Various Types of Traffic (1978)

Traffic Type	Number of Terminals	
	CONUS	Overseas
Voice	200,000	100,000
Interactive Data	13,500	2,500
Computer Data	2,250	250
Narrative	6,000	4,000
Facsimile	3,300	1,200

Local access area traffic destined for other areas uses the military long haul transmission facililies. Some characteristics of current and projected AUTOVON, AUTODIN, and AUTOSEVOCOM systems are given in Table 3. This information was compiled from a number of sources including Rosner (1973), Ochiogrosso et al. (1977), and Levine (1976).

The volume, data rate, duration and delivery delay for traffic generated by various types of terminals is indicated in Table 4. Only nominal values are shown in this table for most of the parameters. Actual values may depart considerably from these nominal values. To be used, data rates must conform to the DOD and FTSC standard rates for synchronous transmission (FTSC, 1973). The standards basically establish rates at increments of 75×2^k bit per second for $k = 1, 2, \ldots, 7$; or $8000 \times k$ bits per second for $k = 1, 2, 7$; plus several exceptions, such as 19.2 and 50 Kb/s. Thus the data rate generated by any given terminal may vary over a fairly wide range. The voice digitization rate depends on the process employed. At the higher rates the voice quality is generally higher for a given bit error rate over the channel.

It is apparent from Table 4 that traffic can be classified according to various characteristics. Three classifications based on channel occupancy are generally used; Class I for continuous traffic, Class II for burst traffic, and Class III for interruptable traffic. Coviello and Vena (1975) characterized each class by examining a variety of terminals including man/man, man/machine, and machine/machine operation. Their results and some others are summarized in Table 5. The distinction, other than occupancy, between Class III and Classes I and II is one of delivery time and service level, rather than any real technical consideration. Thus, although Class I and Class II traffic are handled differently in a network using circuit and packet switching, the Class III traffic can normally be handled by either type of switch depending on specific situation.

2.3 The System Designer

In order to offer a particular service, the telecommunications network must have sufficient functional capabilities or attributes. These functional attributes characterize the three basic elements of the network namely the terminals, the links and the nodes as shown in Figure 8. The terminals perform media conversion functions by transforming source information into the electrical signals used for transferring information. The links provide the transmission path over which the information will flow and the nodes perform various communication processing

Table 3. Current and Projected Characteristics of DCS Backbone Networks

Network	Characteristic	Current (1978)	Projected (1985/1990)
AUTOVON	No. of Subscribers	3.0×10^5	1.0×10^6
	Call Attempts/Day	7.5×10^5	Unknown
	Switch Type	Circuit	Hybrid
	No. of Nodes		
	CONUS	65	43
	Overseas	16	20
	Access Areas	1,500	1,500
	Traffic (Erlangs)		
	Backbone (CONUS)	2,200	2,500
	Node	34	58
AUTODIN	Computers	250	2,500
	Remote Terminal	1,400	25,000
	Messages/Day	3.5×10^5	1.2×10^7
	Switch Type	Message	Packet
	No. of Nodes		
	CONUS	9	24
	Overseas	8	7
	Backbone Traffic		
	Average (b/h)	7×10^9	1.4×10^{10}
	Peak (b/s)	8×10^6	1.0×10^8
	Nodal Traffic		
	Average (b/s)	2.5×10^5	5.0×10^5
	Peak (b/s)	1.0×10^6	1.25×10^6
AUTOSEVOCOM	Subscribers	1,400	10,000
	Terminal Rates		
	Wideband	50 Kb/s	16 Kb/s
	Narrowband	2.4 to 9.6 Kb/s	2.4 to 9.6 Kb/s
	Nodes	AUTOVON	20
	Traffic	Unknown	Unknown

Table 4. Characteristics of Traffic Generated by Various Types of Terminals

	Volume	Digital Rate (one-way)	Call Duration	Delivery Delay
Voice				
PCM	continuous bits	64 Kb/s	minutes	<200 ms
CVSD	continuous bits	16 Kb/s	minutes	<200 ms
LPC	continuous bits	2.4 Kb/s	minutes	<200 ms
Data				
Data Base Update	10^2 b/message	2.4-16 Kb/s	seconds	seconds to minutes
Interactive	10^3 b/message	150 b/s-56 Kb/s	hours (Bursts in seconds)	seconds
Query/Response	10^4 bits/ transaction	150 b/s-9.6 Kb/s	seconds to minutes	<1 second
Bulk	10^5-10^8 bits/ transaction	100 Kb/s	minutes to hours	minutes to hours
Narrative				
Alpha Coded Text	3 x 10^4 bits/page	75 b/s-9.6 Kb/s	seconds to minutes	minutes to hours
Text Editing	10^3 bits/page	75 b/s-9.6 Kb/s	seconds to minutes	seconds
Facsimile				
No Gray Scale	3 x 10^5 bits/page	4.8 Kb/s	minutes	minutes to hours
Half Tone Photo	3 x 10^6 bits/page	9.6 Kb/s	minutes	minutes to hours
Color	10^7 bits/page	1.5 Mb/s	minutes	minutes to hours
Video				
Picture Phone	continuous	6.3 Mb/s	minutes	<200 ms
Color TV	continuous	30 Mb/s	hours	seconds
Slow Scan TV	continuous	100 Kb/s	minutes	minutes

Table 5. Traffic Classifications Based on Channel Occupancy

CHANNEL OCCUPANCY	I CONTINUOUS		II INTERMITTENT		III INTERRUPTABLE	
Type / Characteristics	Broadcast	Dialog	Interactive	Question & Answer	Narrative/ Record	Bulk
Info Transferred Per Transaction	Continuous	Nearly Continuous	<10 to 10^2 Bits	10^2 to 10^4 Bits	10^2 to 10^5 Bits	10^5 to 10^6 Bits
Exchange Mode	Unidirection	Equal Bidirection	Unequal Bidirection	Unequal Bidirection	Primarily Unidirection	Primarily Unidirection
Delivery or Response Time	Real Time	Real Time <0.25 s	Near Real Time <1.0 s	Seconds to Minutes	Minutes to Hours	Minutes to Hours
Originations or Arrivals	Continuous	Few Per Hour	Several Per Hour	Several Per Hour	Few Per Hour	Few Per Day
Duration	Hours	Minutes	Seconds	Minutes	Seconds to Minutes	Minutes to Hours
Connect Time*	N/A	15 to 10 s	Few Seconds	Several Seconds	Minutes	Minutes to Hours
Cross Network Delay	Fixed Fractions	Fixed Fractions	Variable <s	Variable s-m	Long m-h	Long m-h
Call Acceptance	N/A	May be Blocked	Always if Destination Available	Always if Destination Available	May be Delivery Delayed	Always Delivery Delayed
Examples	Radio & TV Broadcasts	Telephone Videophone	Data Processing Alarm/Status Monitoring Telemetry	Data Bank Query	Teletype Facsimile	File Transfer Data Collection Data Distribution

*Depends on Criticality

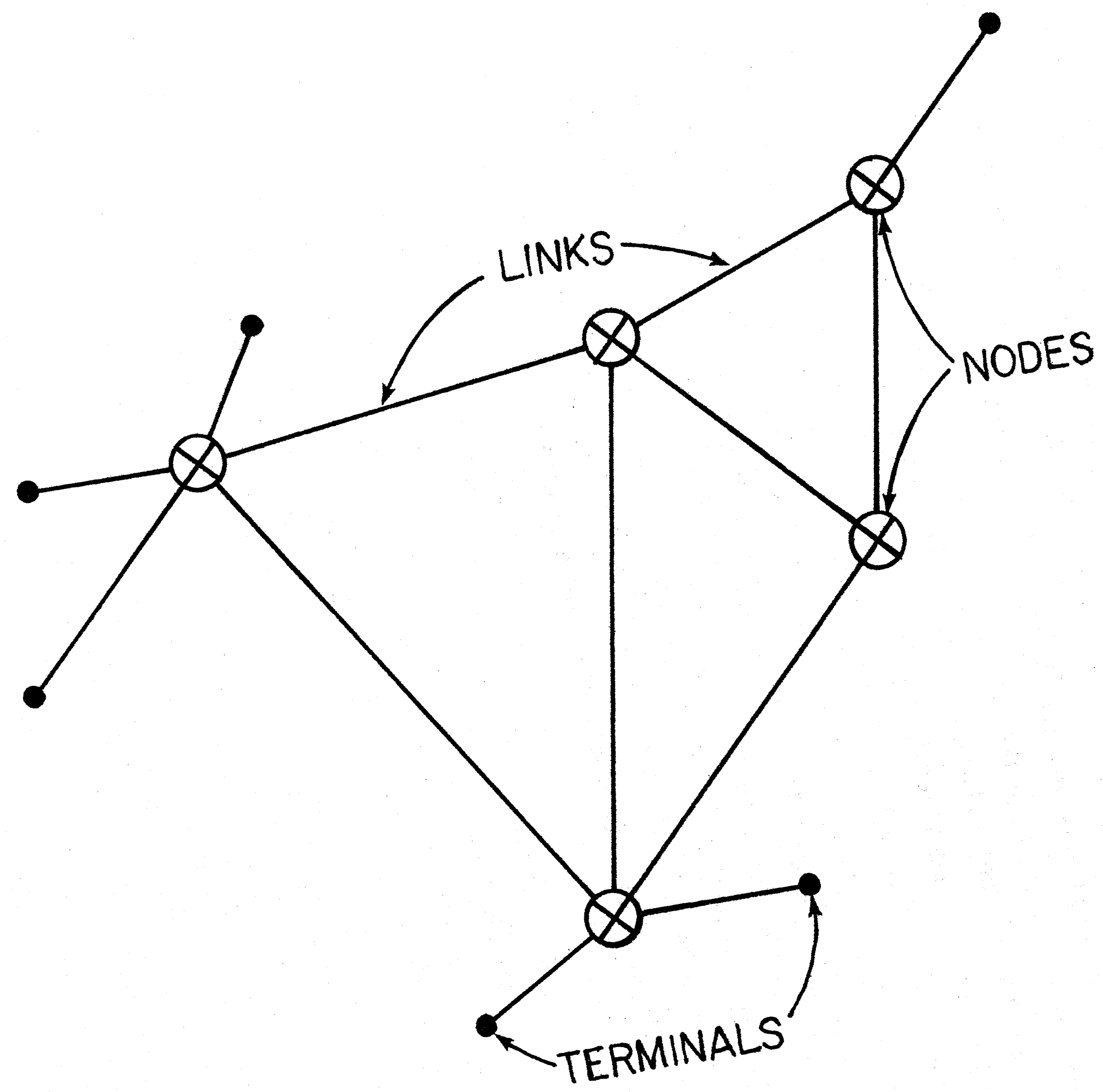


Figure 8. A telecommunications network.

functions such as switching, concentration, and all the functions needed to insure efficient, reliable, and timely transfer of information.

Functional attributes of the network may be implemented in various ways and therefore are not an inherent property of the service offered. For example, the network topology may be a star, loop, or tree; the nodes may incorporate circuit switching, store-and-forward switching, or no switching at all; the transmission facilities may be analog, quasi-analog, or digital.

Figure 9 classifies the terminals in terms of three input and output characteristics as viewed from the system designer's standpoint. Tables 6, 7, and 8 illustrate one method for characterizing various conceptual implementations of networks, switching nodes, and transmission facilities.

Hierarchical structures have been employed in the engineering design of numerous telecommunication networks. At each level of the hierarchy different node and link functions may be specified to meet the overall network design objectives. An example of one hierarchical structure which could be used to access an office complex on a military base to the DCS backbone is shown in Figure 10. In this figure a star connection is employed to connect office terminals to the PABX. A line network is shown to connect several PABX's to the base central office. Several central offices are connected together using a grid network. Different terminals and switch types may be used at each level in the hierarchy. Tables 6, 7, and 8 provide the AUTODIN analyst with blank rows for hierarchical level specification.

2.4 Service and Mission Class Selections

In Section 2.1, the service and mission classifications were based on user-perceived differences in functional characteristics and performance. This approach tends to group the services into categories where similar parameters apply. At the same time it groups users with a common mission into categories where similar values for these parameters also apply.

Ultimately we expect that user parameters and values will be developed for most of the service and mission categories shown earlier in Figures 6 and 7. Initially, however, we have selected one analog service and one digital service for Phase A. Referring to Figure 6 the selected services are as follows:

Classification	Service	
Nature of Signal:	Analog	Digital
Information Form:	Audio	Operator
Type of Interaction:	Human - Human	Human - Machine
Information Flow:	Bidirectional	Bidirectional
Number of Users:	Many - Many	One - One

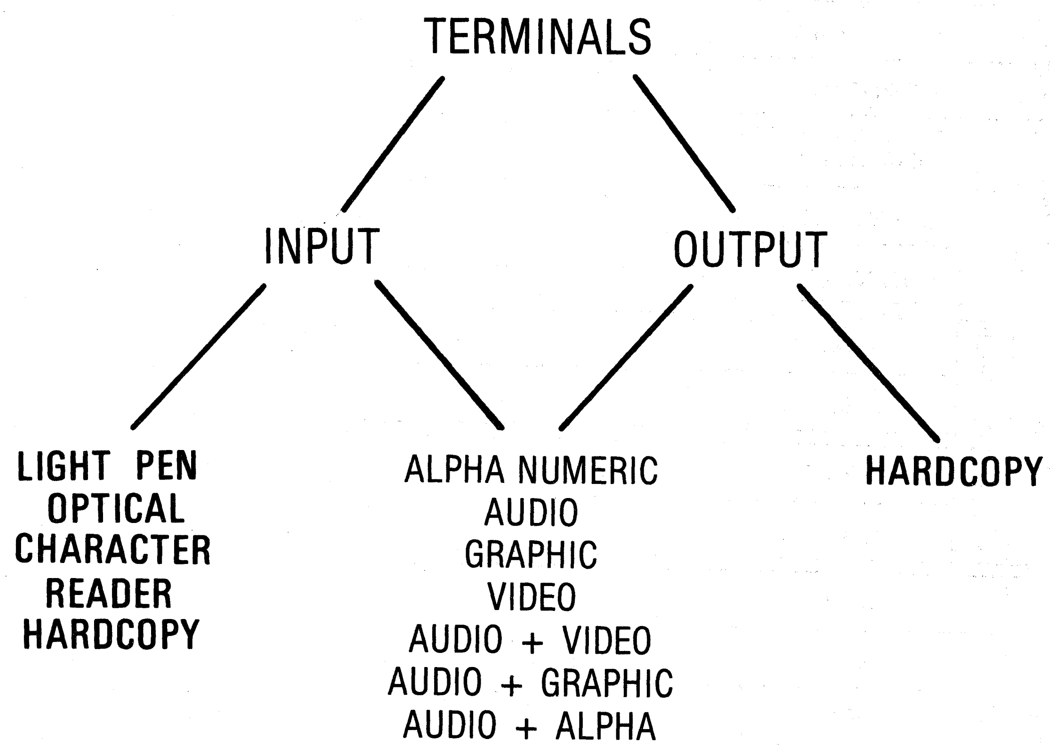


Figure 9. Terminal classifications based on input and output characteristics.

Table 6. Classifications Based on Network Concepts

NETWORK ATTRIBUTES / TYPE & LEVEL	USERS		TOPOLOGY						LINKS				CONTROL			CALL ACCEPTANCE			ROUTING		
									SWITCHED		DEDICATED										
	COMMON	SPECIAL	LINE	TREE	STAR	LOOP	GRID	HYBRID	PT. TO PT.	MULTIPOINT	PT. TO PT.	MULTIPOINT	CENTRALIZED	DECENTRALIZED	DISTRIBUTED	BLOCKABLE	ALWAYS ACCEP.	ACCEPT. BUT THROTTLED	FIXED	ADAPTIVE	SATURATION

Table 7. Classifications Based on Switching Concepts

SWITCHING CONCEPTS / TYPE & LEVEL	CIRCUIT SWITCHED											STORE AND FORWARD SWITCHED										NONSWITCHED		
	CROSS POINTS		ANALOG					DIGITAL				MESSAGE SERVICE		PACKETIZED SERVICES			LINK CONTROL		ERROR CONTROL			ACCESS		
			MATRIX			SIGNALING		MATRIX		SIGNALING														
	METALLIC	NONMETALLIC	FREQ. DIVIDED	TIME DIVIDED	SPACE DIVIDED	PER CHANNEL	COMMON CHANNEL	TIME DIVIDED	SPACE DIVIDED	BORROWED BITS	ALLOCATED BITS	MANUAL	AUTOMATIC	VIRTUAL	DATAGRAM	TERMINAL EMULATED	BITS ORIENTED	CHAR. ORIENTED	NONE	ARQ	FEC	POLLED	RESERVED	RANDOM ACCESS

Table 8. Classifications Based on Transmission Concepts

TRANSMISSION CONCEPTS / TYPE & LEVEL	SERVICE				SIGNAL										MODE						
	FIXED		MOBILE		AMPL.		TIME		CHANNEL-IZATION		FORMAT				OPERATING			PARALLEL		SERIAL	
	PT. TO PT.	MULTIPOINT	PT. TO PT.	MULTIPOINT	CONTINUOUS	DISCRETE	CONTINUOUS	SAMPLED	SINGLE	MULTIPLE	CODED	UNCODED	MODULATED	UNMODULATED	SIMPLEX	HALF DUPLEX	FULL DUPLEX	SYNC	ASYNC	SYNC	ASYNC

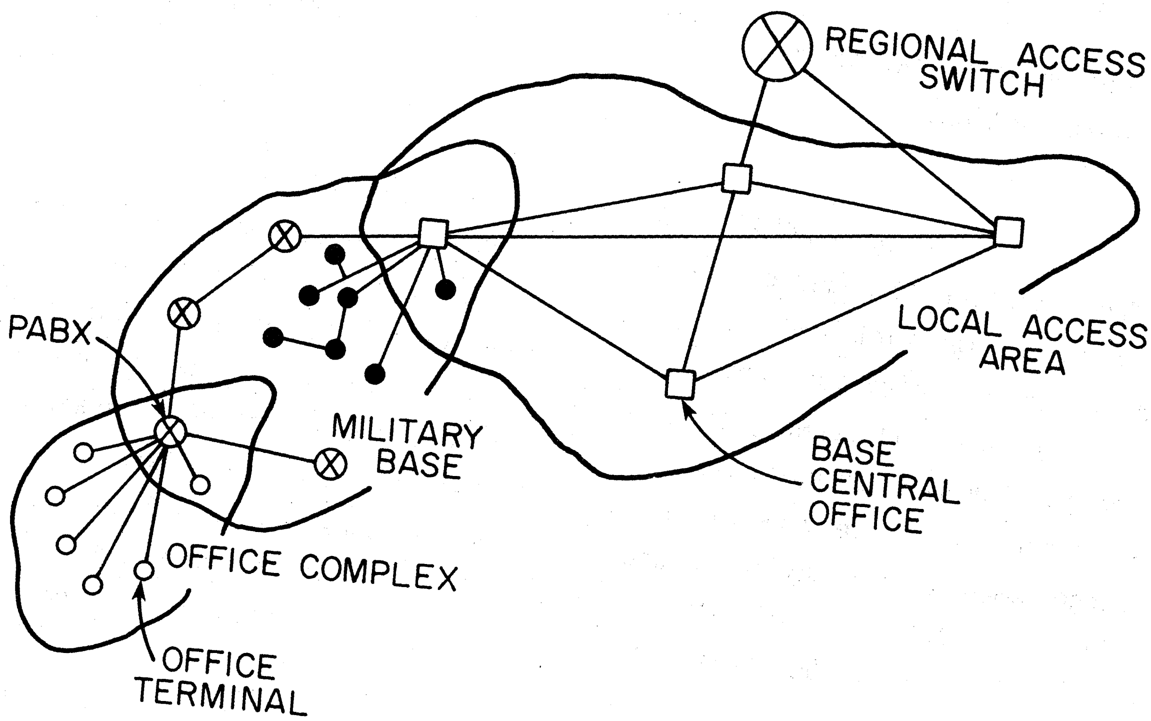


Figure 10. Hierarchical network configuration for access to local and regional switching centers.

Referring to Figure 7 the mission categories selected for these services are as follows:

Mission	Analog Service	Digital Service
Category:	Strategic	Strategic
Information Type:	Command & Control	Operational
Criticality:	Critical	Priority
Classification:	Secure	Nonsecure

The selection of operating modes and the characterization of each mode is discussed in the next section.

3. CHARACTERIZATION OF OPERATING MODES

The selection of modes of operation for the analog service and for the digital service is an important step in the project since the mode of operation identifies a particular communication service. The selection of user oriented parameters and their values will depend, to a large extent, on the functional characteristics assumed for the mode of operation. The ideal complete set of user parameters selected for a given service class would be independent of the system. However, the large complete set may be reduced to a smaller subset when a specific mode of operation is addressed. The primary impact of user-oriented parameters is on users and on service. Secondary impact affects system functions, system design, and system cost.

The mode of operation selected has even greater significance when establishing a relationship between the system-independent user-parameters and system dependent engineering parameters. Measured changes in user parameters indicate when the system performance is changing. Measured changes in the engineering parameters indicate why the system performance is changing.

In the following subsections we describe the methodology employed to select the modes of operation for subsequent analysis.

3.1 Basis for Selecting Modes of Operation

During the initial study phase five criteria were developed as the basis for selecting the operating modes. These are: 1) anticipated demand; 2) impact on military operations; 3) prospects for future implementation; 4) most stringent requirements on system design; and 5) data base availability. The fifth was included because the availability of data affects how well the user-oriented parameters and values can be translated into engineering parameters and values. Subsequent discussions with the Defense Communications Engineering Center (DCEC) provided guidance concerning which of the above criteria should be given the greatest weight. The principal criteria guiding the selection were identified as follows:

a. The mode selected should be more vulnerable than other modes to transmission impairments. In this way, transmission facilities designed to support the selected mode would provide satisfactory service for other modes.

b. The mode selected should have relatively high utilization level and should involve high priority traffic.

Thus the greatest weight was given to items 1), the anticipated demand and 4), design requirements which put more strain on the transmission facilities.

There are numerous causes of service impairments in any telecommunications system which could also affect the mode selection process. These impairment causes may result from exterior events or interior events as illustrated in Figure 11.

For our purposes the system impairments are the primary interest. These may occur at the terminals, in the nodal elements or on the transmission links as shown in Figure 11. Eighteen leading impairments resulting from these causes are shown in Figure 12 for telephone networks. These were derived from many references. See, for example, AT&T (1977). A simplified diagram of a typical telephone network is shown at the top of Figure 12. In this diagram a conventional telephone is homed on an analog switch via a two-wire loop. An analog trunk connects to another analog switch. Subsequent digital trunking and switching requires conversion from analog to digital and back to analog at the destination telephone. The eighteen impairments are listed on the side of Figure 12 and the arrows indicate the various points in the system where they might occur. It is important to note that most of the impairments listed can occur on the transmission links, including the access loops and the trunks.

The three leading impairments which apparently occurred most often in the 1970's are indicated in Figure 13. These are Gaussian noise, echoes, and signal loss. Echos are not related to the transmission links but to end interfaces or terminals themselves.

These eighteen impairments can also be extended to the AUTODIN II network as shown in Figure 14. A simplified digital service mode of operation between

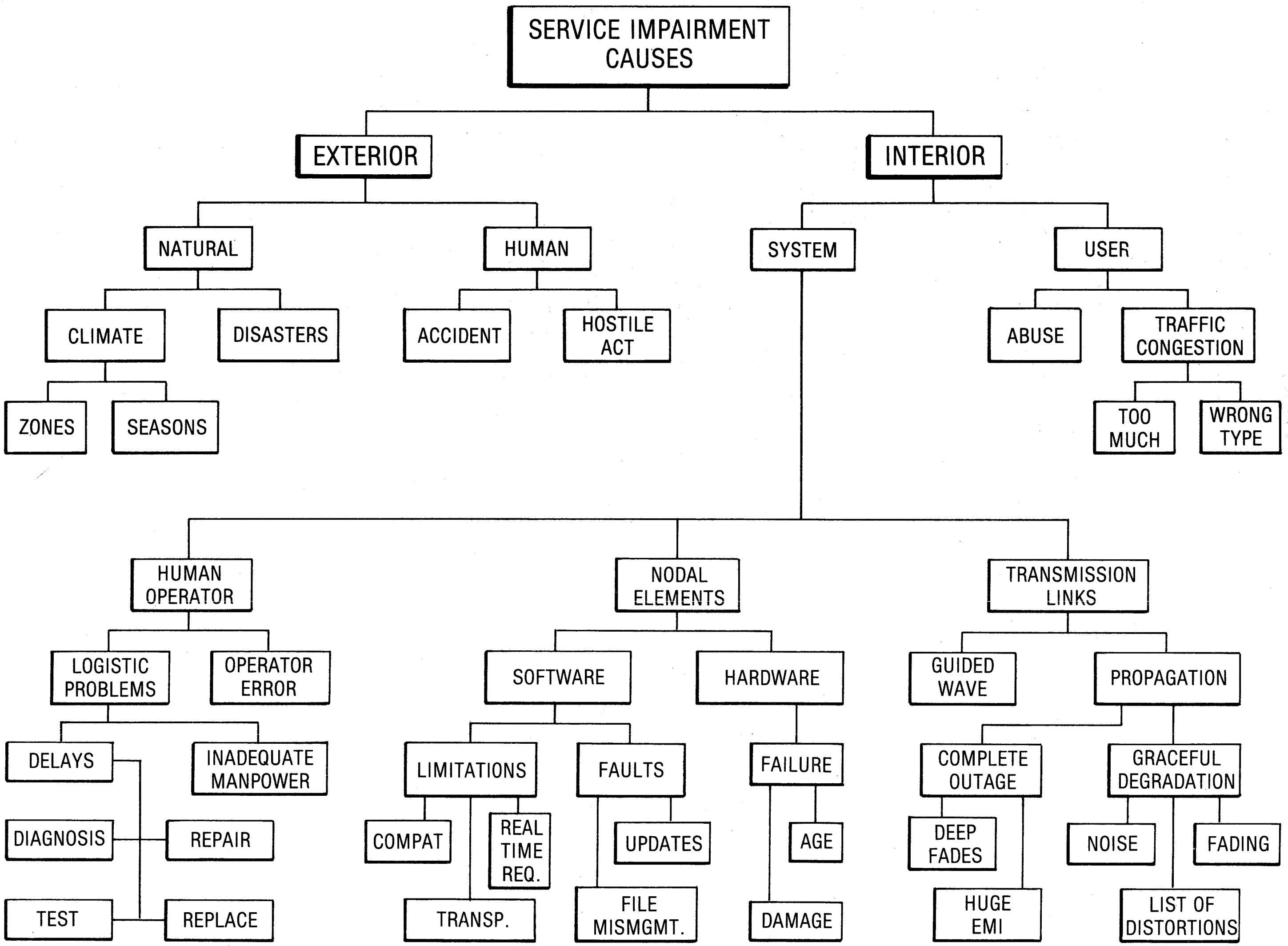
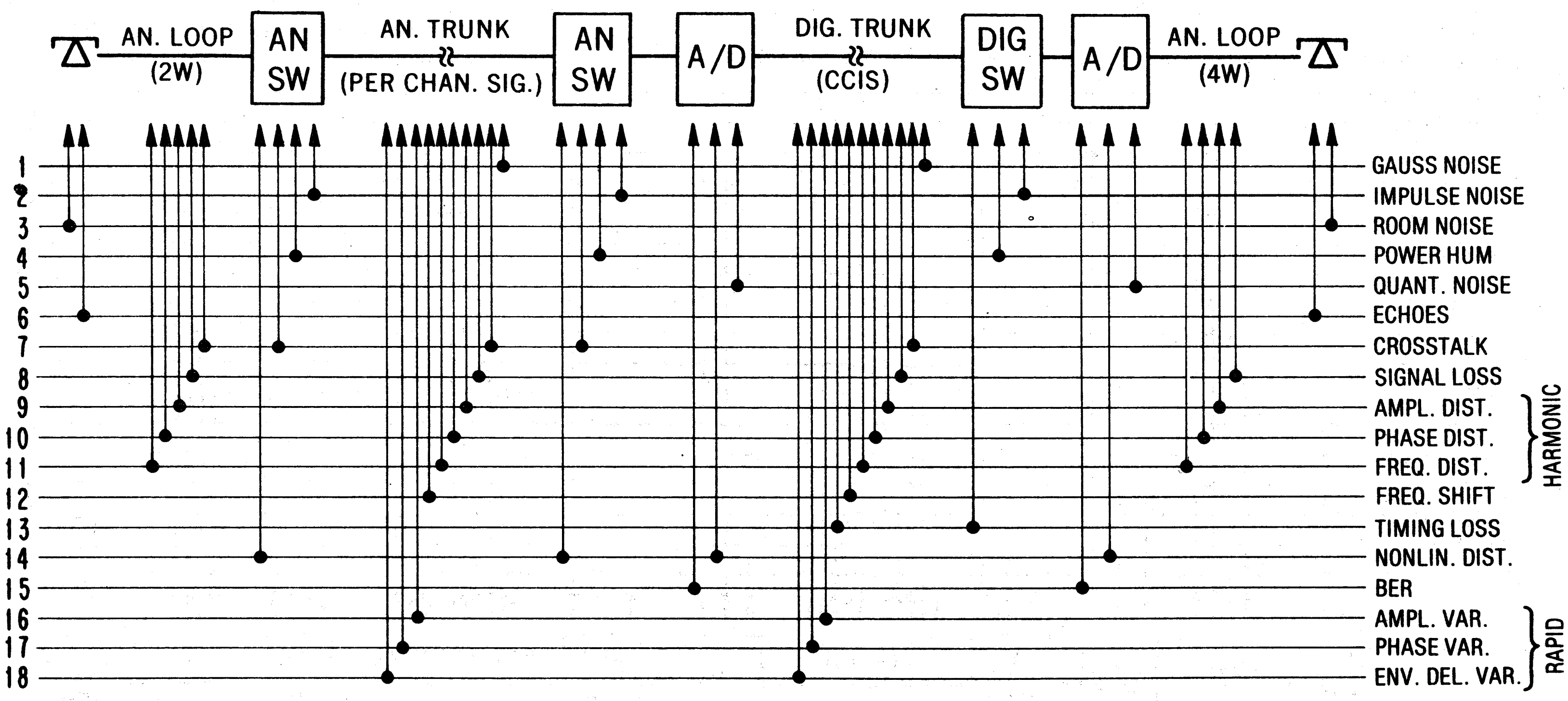
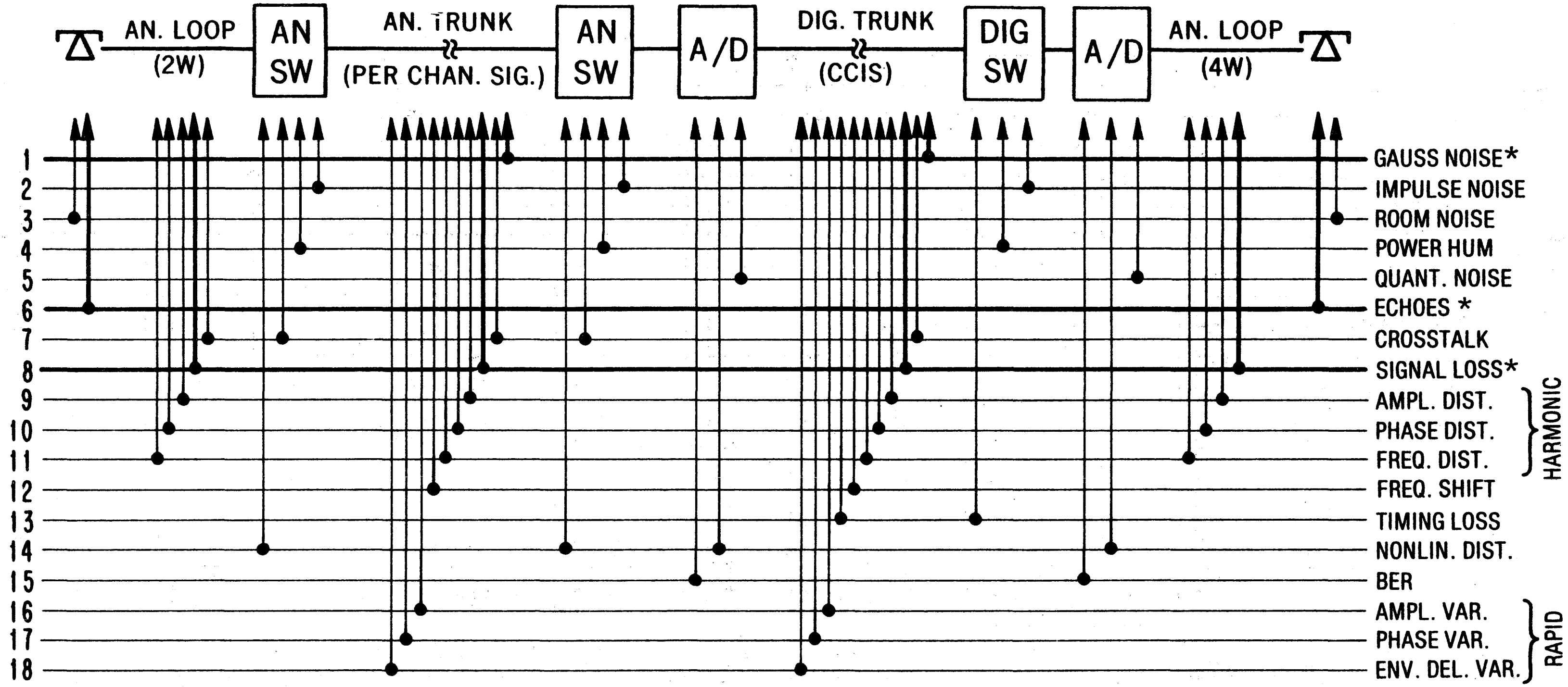


Figure 11. Examples of interior and exterior causes of service impairments.



AN - Analog
SW - Switch
CCIS - Common Channel Interswitch Signaling

Figure 12. Eighteen leading impairments to telephone networks, circa 1970.



*Three leading impairments.

Figure 13. Three leading impairments for analog services, circa 1970.

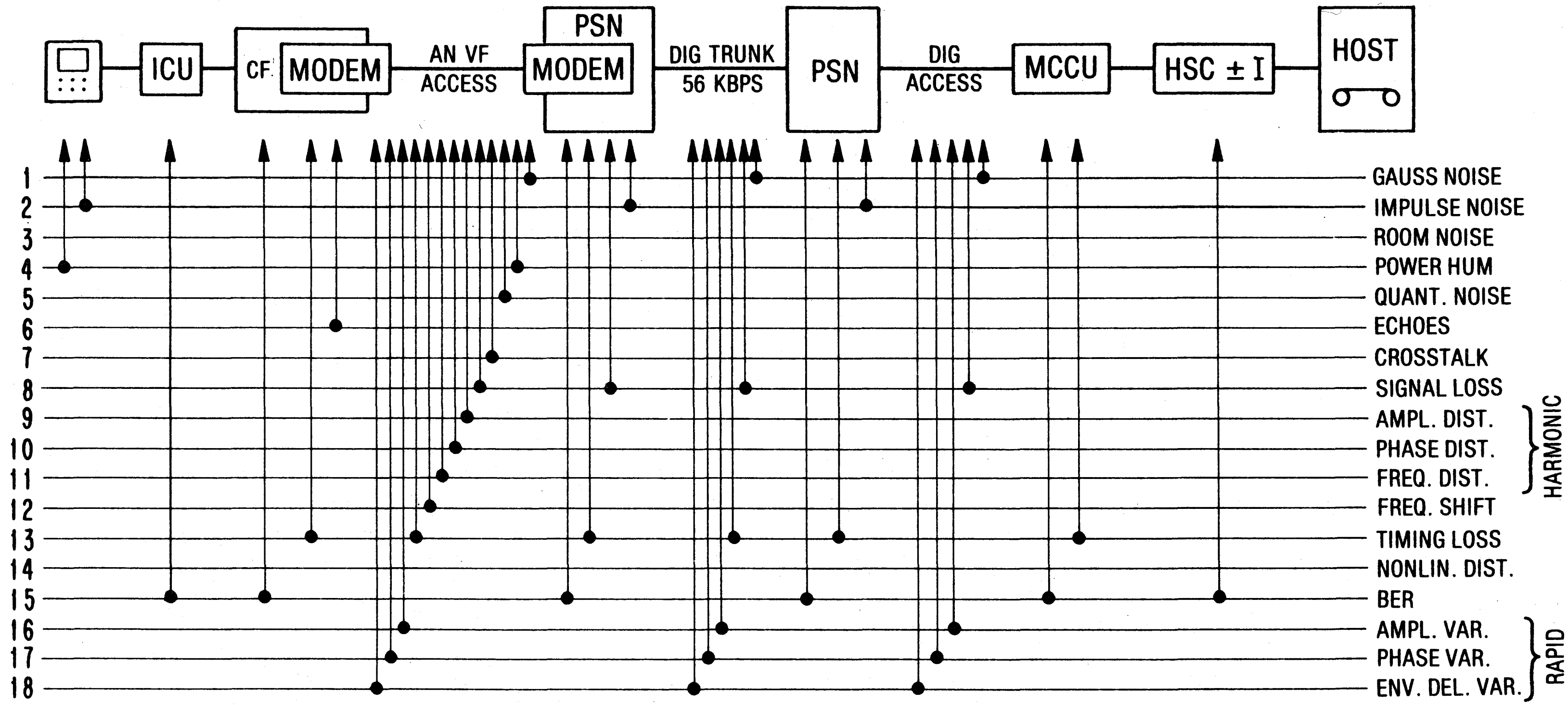


Figure 14. Eighteen leading telephony impairments extended to AUTODIN II digital service.

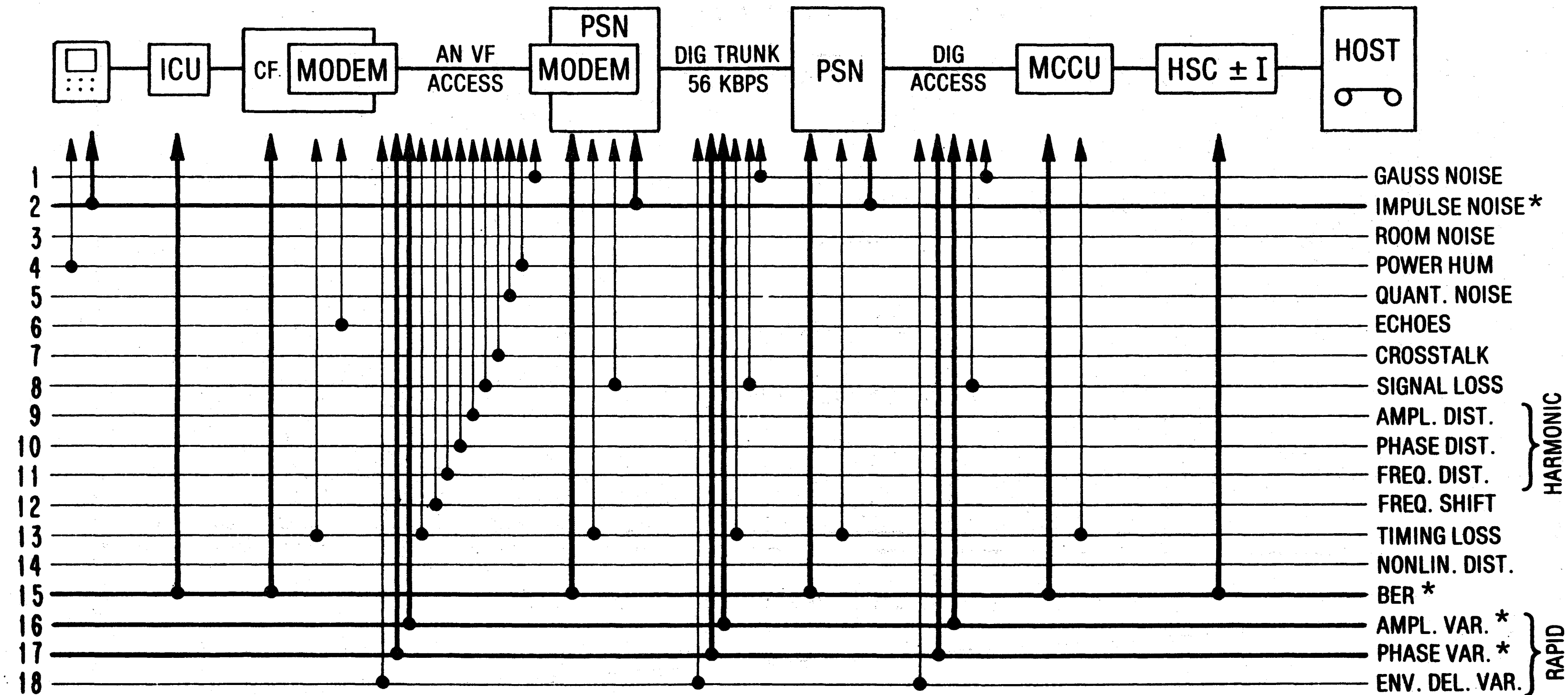
a terminal operator and a host computer is depicted across the top of this figure. The operator's terminal on the left is shown connected to a host computer on the right. The blocks indicate the various network units involved. From left to right they are, the Interface Control Unit (ICU), the Communications Facility (CF) which includes a modulator-demodulator (modem), the analog voice frequency (VF) access line, the Packet Switch Node (PSN), a 56 Kb/s digital trunk to a second PSN, and finally a digital access line to a Multiple Channel Control Unit (MCCU), and Host Specific Control and Interface (HSC&I) to the host computer. Again the arrows indicate where each of the eighteen leading impairments might occur. The four leading impairments for this digital case are shown in Figure 15. In this case rapid amplitude and phase variations on the transmission links cause higher binary error rates (BER) at the nodes.

Different transaction profiles apply to the network elements shown in Figure 12 for telephone service and in Figure 14 for the AUTODIN II data service. For these profiles the functions of a communications session may be divided into three phases: the access phase, the transfer phase, and the disengagement phase. Each phase requires a finite period of time as depicted in Figure 16 for the telephone service mode. Time increases toward the bottom of the figure. A transaction profile for the access phase of the telephone transaction is illustrated step by step in Figure 17. In this profile it is assumed that successful access is achieved.

The vertical axis of the profile represents elapsed time increasing from top to bottom. The circles indicate when and where system specific events occur in the system block diagram.

Figure 18 illustrates a successful disengagement transaction for the same telephone network. A profile for one digital service mode on AUTODIN II is given in Figure 19 for an assumed successful access.

More specific profiles are not essential to the initial development of user parameters and values because these parameters and values are supposed, with certain qualifications, to be system independent (i.e., they are not intended to depend on any particular equipment, protocols, or network architectures). The main qualifications are practicality and cost. For example, a users' requirement for extremely short access time to another user may be much less than the access time achievable on any switched network. A dedicated line could meet the access time requirement, but at a considerable increase in cost. User parameter values must be determined by evaluating user effectiveness in accomplishing a mission,



*Four leading impairments.

Figure 15. The four leading impairments for digital services, circa 1970.

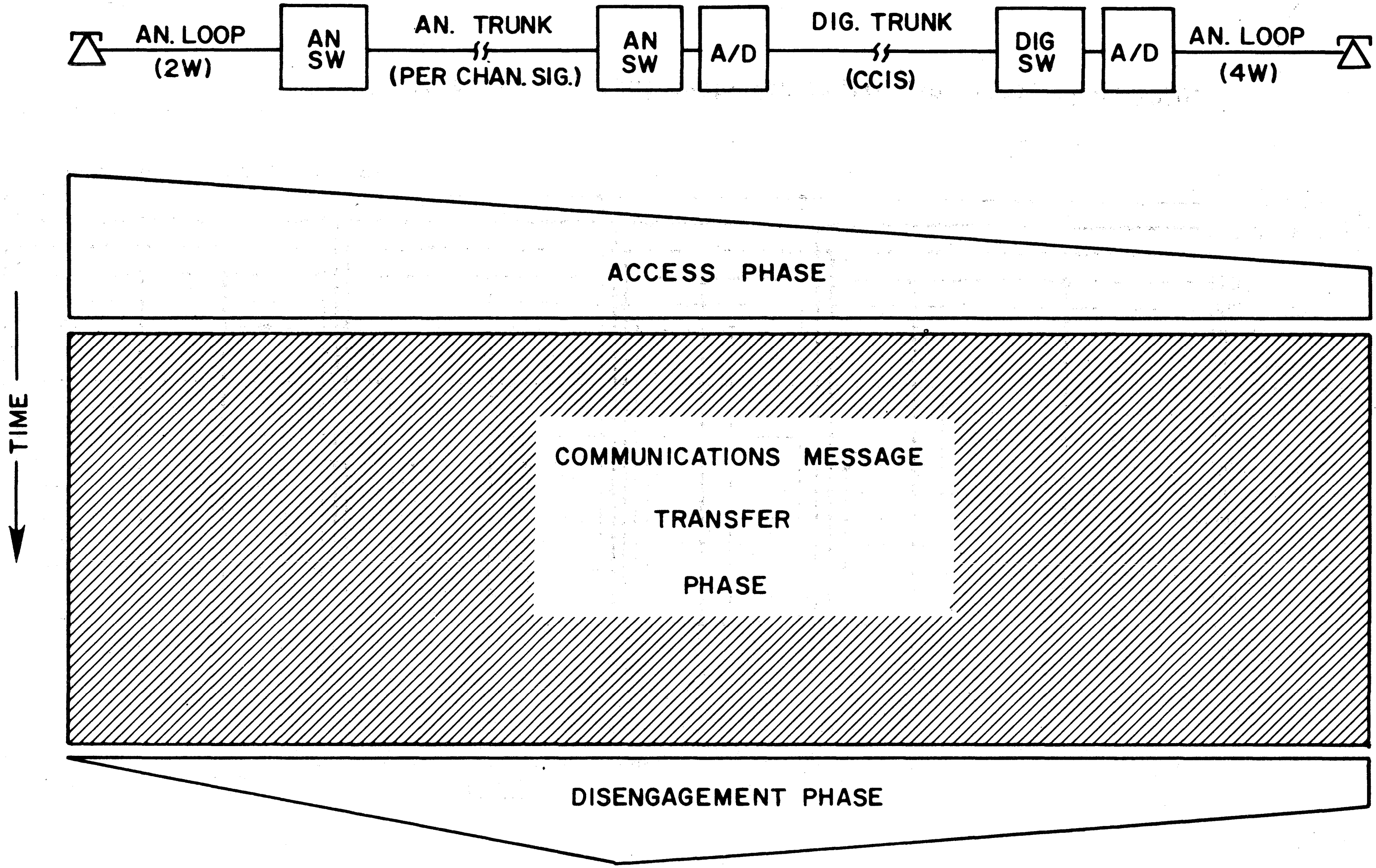


Figure 16. Functional phases of a telephone session.

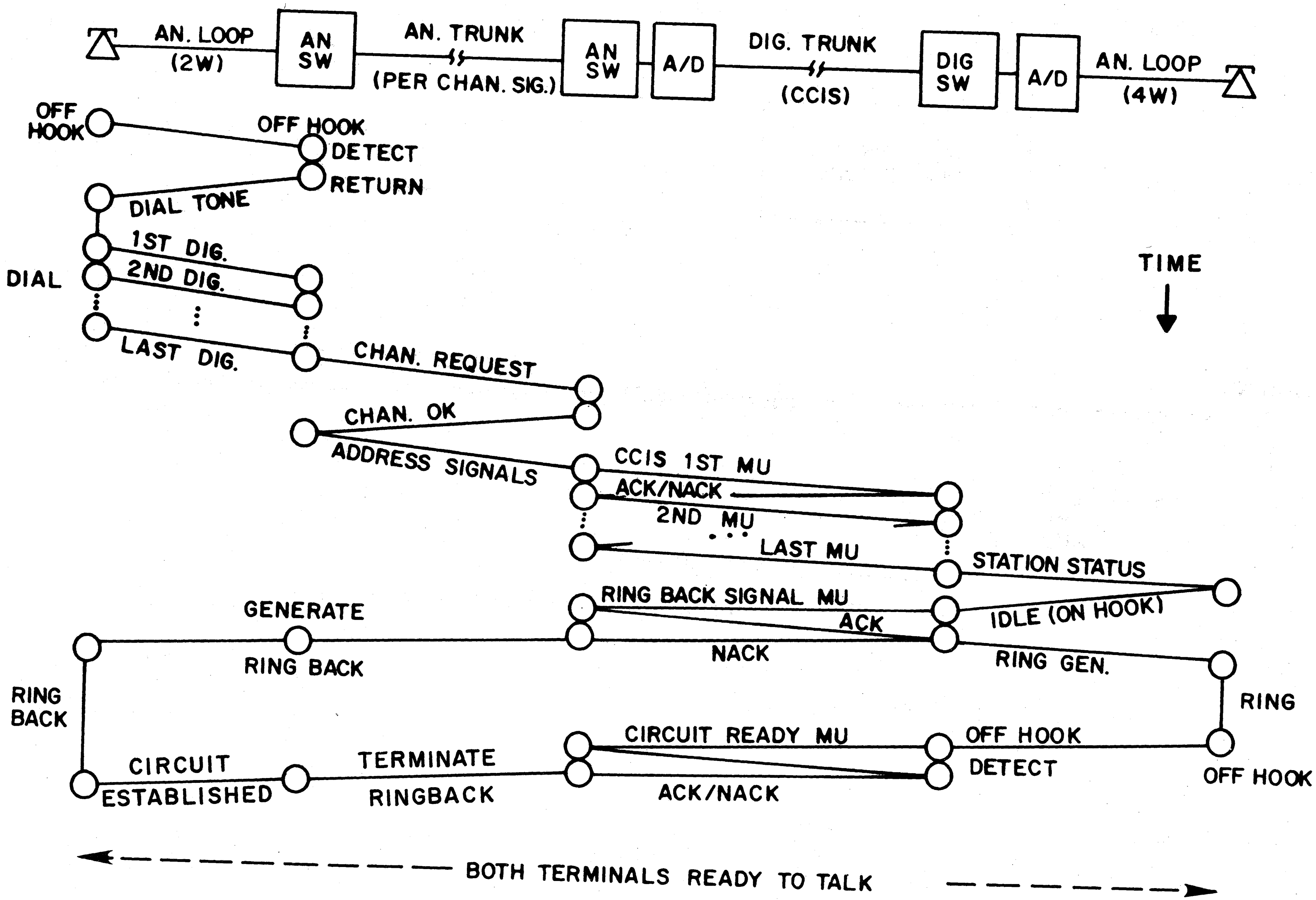


Figure 17. Transaction profile assuming successful access for analog service mode.

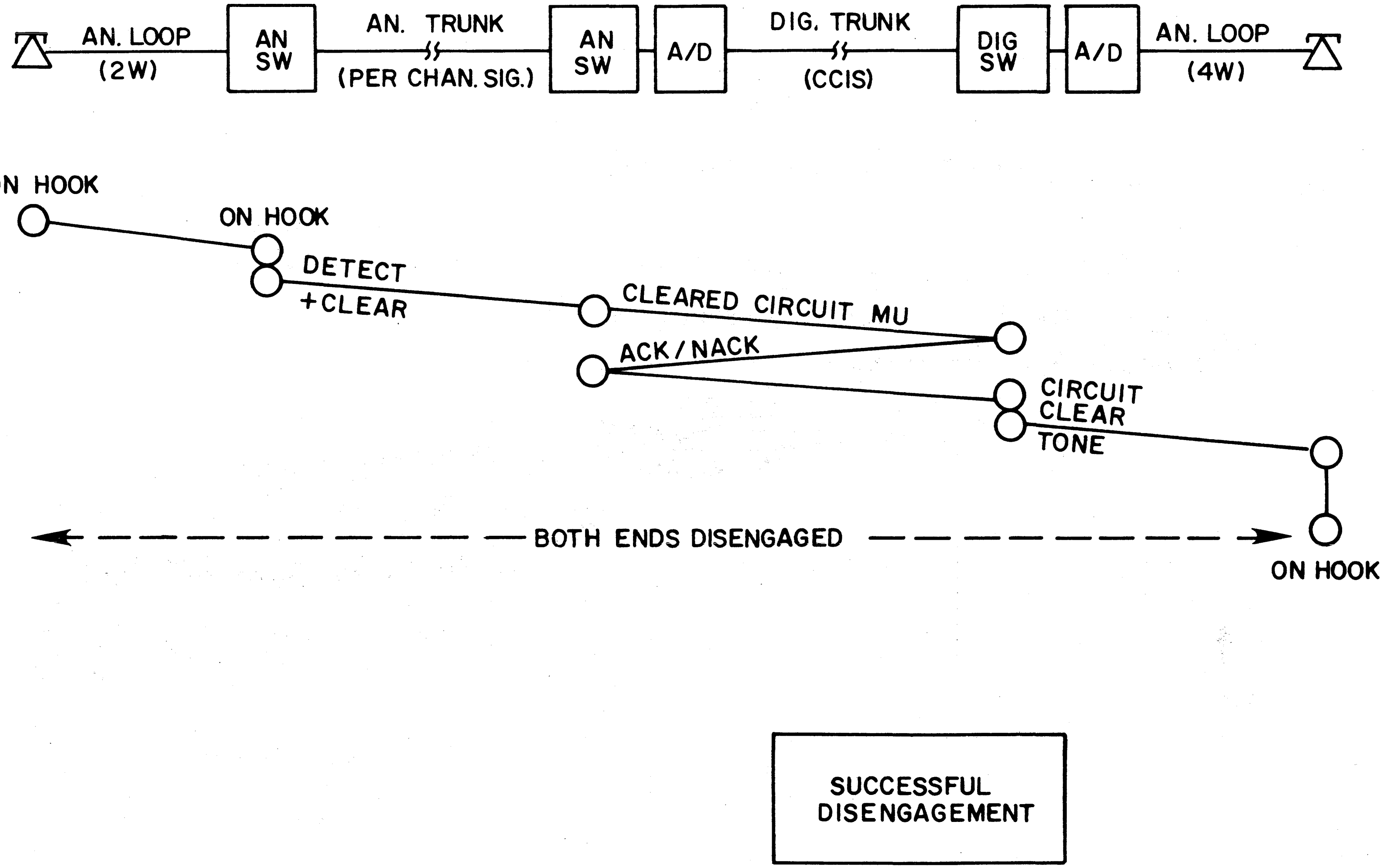


Figure 18. Transaction profile assuming successful disengagement for analog service mode.

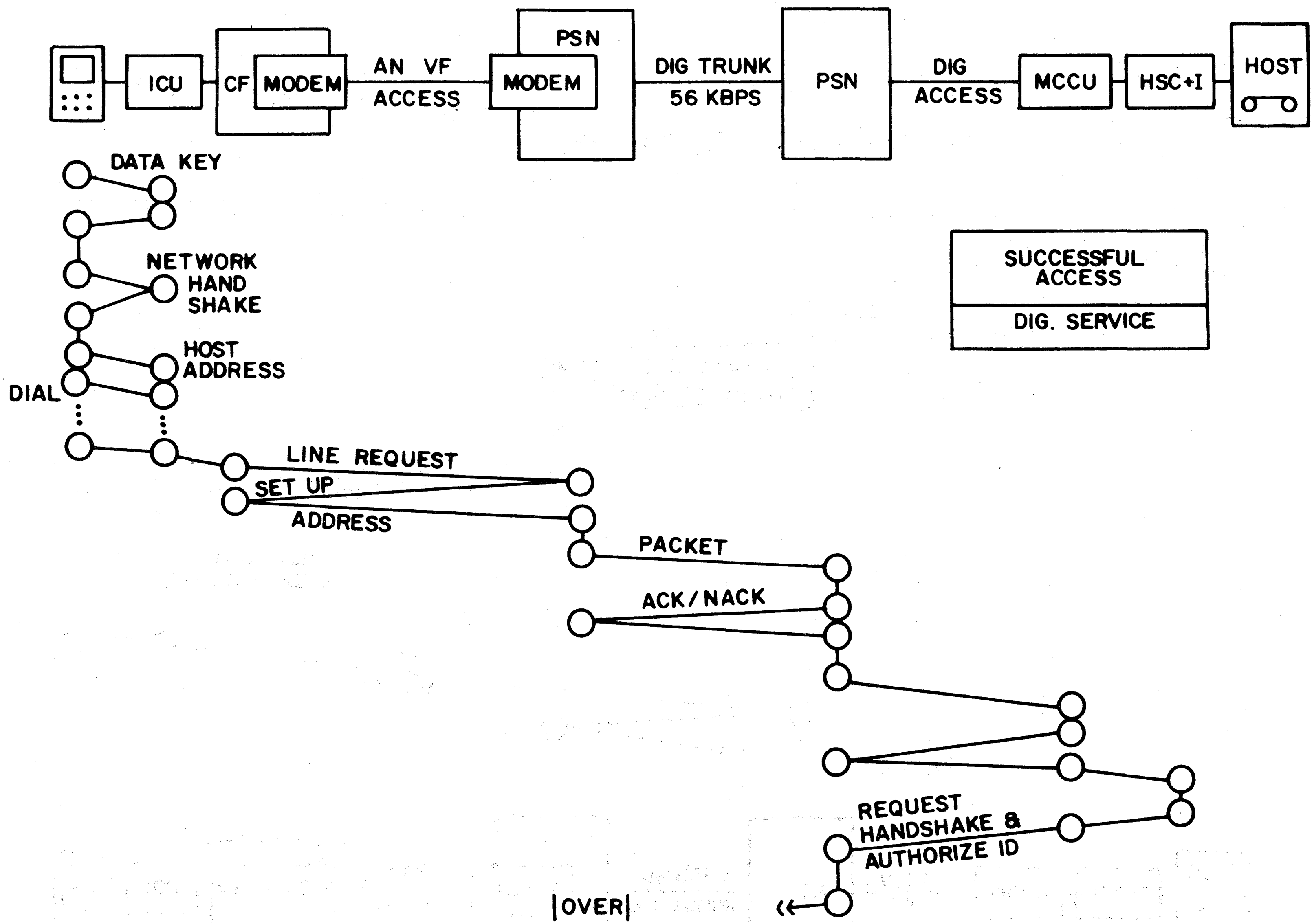


Figure 19. Transaction profile assuming successful access for digital service mode.

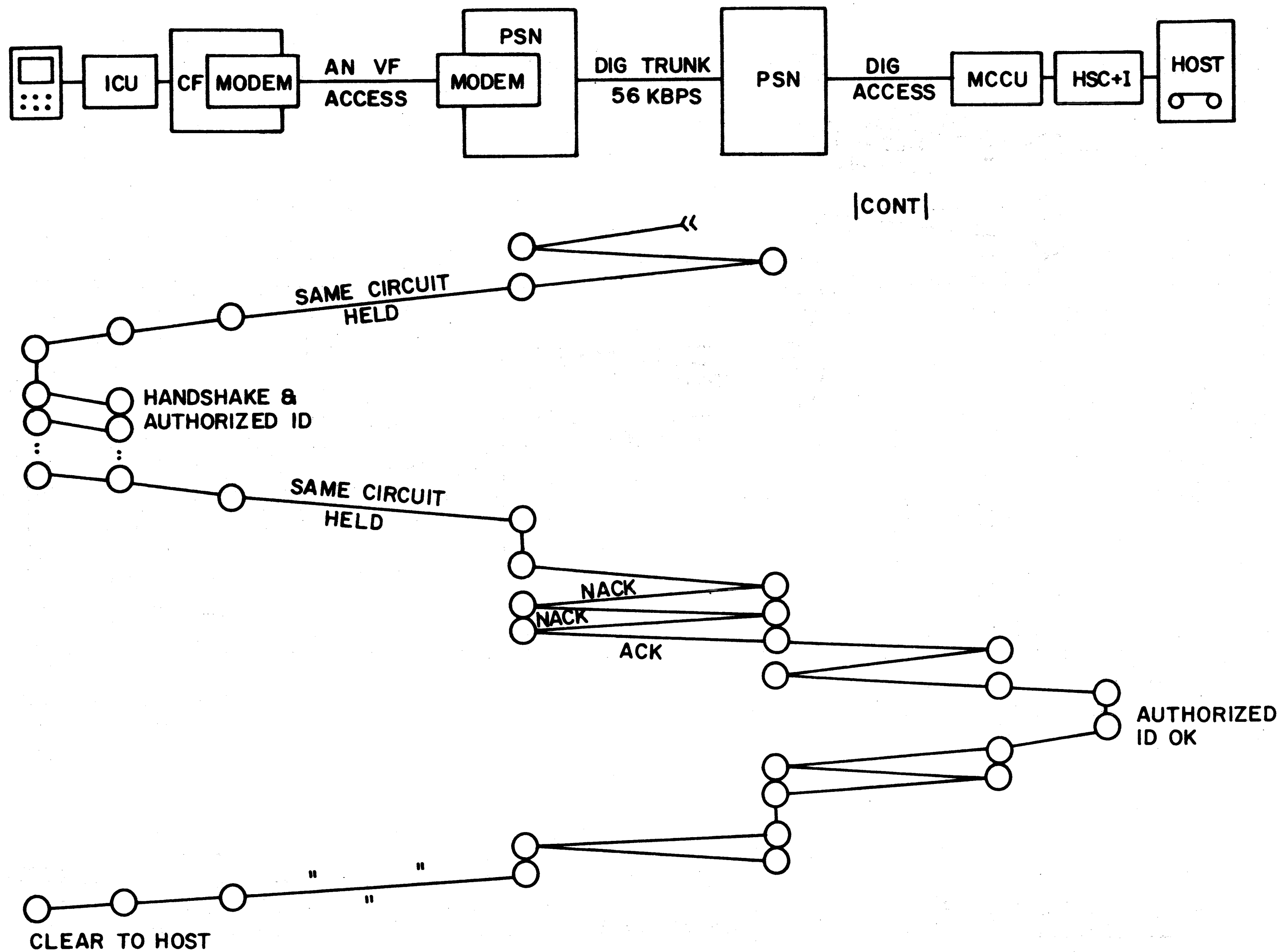


Figure 19. Continued.

relating this to system performance, and then to cost. For this reason the values assigned to user parameters cannot be completely independent of the system.

In the two following subsections two specific modes of operation are selected. Realistic user requirements on service will eventually be translated to choices of practical system alternatives.

3.2 Selection of the Digital Service Mode

This section summarizes the work on the selection of AUTODIN II digital service modes. All modes in question belong to one of four categories.

Mode IB : Synchronous, character oriented, full duplex (FDX), with character and block parity checks for automatic repeat request (ARQ), similar to binary synchronous (BISYNC) operation and the mode used in AUTODIN I. Data rates between 150 b/s to 56 Kb/s.

Mode IIA : Asynchronous, character mode with FDX. No error control coding, but a provision of echo option. ASCII standard. Rates between 110 b/s and 2.4 Kb/s.

Mode IIAH : An asynchronous character mode similar to IIA. Half duplex (HDX) with a second signaling link, but without the echo option.

Mode VI : Synchronous, binary, FDX mode with ADCCP protocol but with 32-bit CRC for error detection. The ADCCP protocol is nearly identical to the Synchronous Data Link Control (SDLC) of IBM and the High-Level Data Link Control (HDLC) of ISO. It has been approved by ANSI. Automatic repeat request (ARQ) is used to correct errors. Data rates range from 1.2 to 56 Kb/s.

The two basic criteria used to select the modes at their more common source data rates are:

(a) Technical service vulnerability to AUTODIN II backbone network impairments.

(b) The service utilization level of each mode subcategory in terms of busy hour traffic volumes and the number of terminal installations.

In what follows, the two approaches (a) and (b) are discussed in detail.

(a) Technical Service Vulnerability

The general approach to mode vulnerability is as structured in Table 9. The first column lists the four modes: IB, IIA, IIAH, and VI. The second column specifies the prevalent source data rates in either b/s or Kb/s. The third, fourth, and fifth columns identify applicable error control methods, if such are used. Thus, character and/or block parity check bits are employed in mode IB for ARQ operation. An ARQ technique with a 32 bit cycle redundancy check (CRC) sequence is used in mode VI. Mode IIA shows the so called echo option, which may or may not be used with each of its six data rates. Mode IIAH involves no error control.

Columns 6 through 12 in Table 9 assess the relative severity of backbone impact on the perceived service. Another way to look at it is as the users' service vulnerability to AUTODIN II network impairments. Clearly, if the main network suffers a total breakdown, no mode can offer a service through it. However, by the same token there is a range of realistic day-to-day variations that either enhances or reduces the technical fidelity of the network operations. Columns 6 through 12 are concerned with just such realistic impairments on the AUTODIN II backbone. The job is to discern how such impairments manifest themselves on the individual service modes quoted earlier. It appears that there are several effects involved here. The key effects are noted either during the communications message transfer phase or during the initial access phase.

During the message transfer phase, individual bit or word errors may pass undetected and uncorrected through some leg of the network. When delivered to the user terminal, such error rate (BER) degrades the service. The BER effects are assessed qualitatively, in column six of Table 9.

Column six contains descriptors, med (for medium) or low, or the space is blank. The word "MED" implies that backbone performance has a noticeable nominal effect on the end-to-end service in terms of BER for the particular mode and data rate. Consider, for instance, the 56 Kb/s modes IB and VI. There the terminal access lines to the AUTODIN II Packet Switch Node (PSN) are likely to be high quality wideband digital lines, such as Type 8856 DDS service terminals. The BER on terminal accesses would be on the same order of magnitude as the BER on backbone links, perhaps both around 10^{-6} on the physical channel without error control. Given ARQ and CRC protection on both the access lines and the backbone links, their BER performances are improved comparably. Thus, backbone effects are apt to be noticed at the service end, but they would not dominate the BER level. This is the reason for entering "MED" in column six, Table 9, for the 56.0 Kb/s mode IB and mode VI.

Table 9. Relative Service Vulnerability

Mode	Source Rate (b/s)	Error Control On Terminal Access			Relative Severity Of Backbone Impact On						
		Char/ Block	Echo Option	CRC	BER	Rate Effic.	Waiting For Access			Service Outage	Other Circum.
							>0s	>1s	>30s		
IB	56.0 K	X			MED	MED	MED	LOW		MED	VOICE
	9.6 K	X			LOW						
	4.8 K	X			LOW			LOW			
	2.4 K	X			LOW	LOW	LOW	LOW	LOW	LOW	
	1.2 K	X			LOW	LOW	LOW	LOW	LOW	LOW	MUX
	300	X			LOW				LOW		
	150	X			LOW				LOW	LOW	
IIA	2.4 K		X		MED				LOW		
	2.4 K					LOW	MED	LOW	LOW	LOW	
	1.2 K		X		MED	LOW			LOW	LOW	MUX
	1.2 K					LOW	MED	LOW	LOW	LOW	MUX
	600		X		LOW						
	600					LOW	LOW		LOW		
	300		X		LOW						
	300					LOW	LOW		LOW		
	150		X		LOW						MUX
	150					LOW	LOW		LOW	LOW	MUX
	110		X		LOW	LOW			LOW	LOW	MUX
	110					LOW			LOW	LOW	MUX
IIAH	2.4 K							LOW	LOW		
	1.2 K							LOW	LOW		
	300										
	110										
VI	56.0 K			X	MED	HIGH	MED	MED	LOW	MED	VOICE
	19.2 K			X	LOW	LOW					VOICE
	9.6 K			X	LOW	LOW					
	4.8 K			X	LOW	MED		LOW		LOW	
	2.4 K			X	LOW	MED	LOW	LOW	LOW	LOW	MUX
	1.2 K			X	LOW	MED	LOW	LOW	LOW	MED	MUX

Table 9 shows two other medium BER effects for mode IIA. They are shown to occur for the echo options of 2.4 Kb/s and 1.2 Kb/s asynchronous FDX over voice-lines. Without the echo, such digital streams may suffer error rates as high as 10^{-4} to 10^{-3}. Backbone degradations are then entirely negligible in the BER sense. The echo option, however, may reduce the access line undetected error rate to the rough order of 10^{-6}.

All modes without error control are left blank in the BER column, because the AUTODIN II impairments are expected to be entirely negligible compared to the 10^{-4} to 10^{-3} BER levels expected for the modes in question. Other modes show "low" BER effects. For example, the 9.6 Kb/s and 4.8 Kb/s equalized access links are apt to have binary error rates slightly above the backbone BER, perhaps by a factor of 10.

Column 7 of Table 9 deals with throughput rate efficiency. If placed back-to-back, the source and destination terminals can run at their intended ideal rate (viz., 56.0 Kb/s in the very first row for mode IB). In the actual network deployment, however, where errors are detected and repeats requested, the real throughput of data is lower. If this useful rate reduction is primarily due to access lines outside AUTODIN II, then the backbone has no effect. On the other hand, as the terminal feeders are improved to cause less relative throughput losses, the network ARQ's become more noticeable. In column 7, descriptors such as low, med, and high are used.

The highest rate efficiency vulnerability to the backbone ARQ's is assigned to mode VI, 56.0 Kb/s. The synchronous binary accesses are usually well engineered. Their lengths tend to be short. The backbone, likewise well engineered, contains numerous links, PSN's, and processes for routing and storing. Their total effect on the rate can be high.

Columns 8, 9, and 10 are concerned with the waiting times during the access phase. The three columns depict different waiting time thresholds. For high speed machines, where every wait is a lost resource, the user may be interested in instant access. Every measurable (e.g., >0 second) delay may be of concern to such a customer. Other application types, such as interactive (I/A), query/response (Q/R), narrative, or bulk data transfers, may put emphasis on nonzero thresholds. The >1 second threshold may be near the perception level of a skilled terminal operator. Delays longer than 30 seconds may be unacceptable to even the slowest console operators.

The relative severity of backbone impact on these three delay thresholds is estimated in Table 9. Note that all effects are either medium, low, or blank.

The rationale used is as indicated above. It stresses the visibility of AUTODIN II backbone network impairments to the customer access phase.

Column 11 of Table 9 is concerned with service outages. Because of the expected redundant deployment of AUTODIN II links and modes, the backbone is not expected to be the major contributor to service outages.

The final column of Table 9 is unique. It identifies other or special circumstances under which backbone impairments may become bothersome to users. For instance, digitized voice such as pulse code modulation (PCM) or continuously variable slope delta (CVSD) may be degraded if voice packets were used over a network with large random delays. Other circumstances may involve multiplexing or intelligent concentration of lower speed modes, where blocking losses and delays may occur.

Table 9 suggests that mode VI, 56.0 Kb/s, must be the most vulnerable. After all, no other mode contains the sum of one high, four med's, one low, plus the potential voice circumstances. To render this scoring process more quantitative, one can use various weighted scoring methods. We have used the simple scoring table of Table 10. The maximum possible score is 30, while the minimum is 0.

Table 10 enables a summary scoring of all digital modes/rates listed in Table 9. This final relative scoring is given in Table 11. The relative scores are technical in part only. That is, BER, rate efficiency, access waiting times, etc., are indeed technical in nature. However, the scoring method is quite subjective. It should be used with knowledge of its limitations.

Nevertheless, Table 11 does complete our first approach based on the technical assessment of service vulnerabilities for different digital modes. The most vulnerable mode is VI, 56.0 Kb/s (with 22 points), followed by IB, 56.0 Kb/s (18 points), followed by VI, 1.2 Kb/s (16 points). The second approach is discussed below.

(b) Service Utilization Levels

Defense Communication Agency's SYSTEM PERFORMANCE SPECIFICATION FOR AUTODIN II, PHASE I (DCA, 1975), pp. 216-238, gives a subscriber listing for the projected AUTODIN II network. The scenario lists eight switches: Albany, Andrews, Ft. Detrick, Gentile, Hancock, McClellan, Norton, and Tinker. Of these, Andrews has the most subscriber terminations. It is second behind Tinker in the total traffic throughput, and third behind Gentile and Tinker in PSN connectivity.

To ascertain which digital modes are used the most, we have scrutinized Andrews traffic only. To include the statistics from all AUTODIN II switches, the size of this job would have to be magnified at least five-fold. This was considered unnecessary under the premise that Andrews utilization is typical of the entire AUTODIN II.

Table 10. Relative Scoring Table

	Relative Severity Of Backbone Impact On						
	BER	Rate Effic.	Waiting For Access			Service Outage	Other Circum.
			>0s	>1s	>30s		
HIGH	6	6	3	3	3	6	3*
MED	4	4	2	2	2	4	3*
LOW	2	2	1	1	1	2	3*
(BLANK)	0	0	0	0	0	0	0

*Any item listed under Other Circumstances is given the weight 3.

Table 11. Technical Severity Score of Backbone Impact on AUTODIN II Service Modes

Mode	Source Rate (BPS)	Error Control On Terminal Access			Relative Score
		Char/ Block	Echo Option	CRC	
IB	56.0 K	X			18
	9.6 K	X			2
	4.8 K	X			3
	2.4 K	X			9
	1.2 K	X			12
	300	X			1
	150	X			3
IIA	2.4 K		X		5
	2.4 K				8
	1.2 K		X		12
	1.2 K				11
	600		X		2
	600				4
	300		X		2
	300				4
	150		X		5
	150				9
	110		X		10
	110				8
IIAH	2.4 K				2
	1.2 K				2
	300				0
	110				0
VI	56.0 K			X	22
	19.2 K			X	7
	9.6 K			X	4
	4.8 K			X	9
	2.4 K			X	14
	1.2 K			X	16

Andrews statistics are listed in Table 12. The first two columns give the modes and their data rates, just as in Tables 9 and 11. The third column of Table 12 adds a terminal example to illustrate the nature of the computer, its front end, or teletypewriter installation. Next, in column 4, the numbers of such installations are given. All of these installations are homed to Andrews. However, there are a few instances of double homing to other PSN's.

The fifth and sixth columns show the traffic in kilobits per hour (Kb/h). The traffic is divided into two columns, denoted as transmitted and received. The transmitted traffic is the total generated by the installations that are homed on Andrews. The received traffic is generated elsewhere, but is intended for Andrews installations.

To assist with the interpretation of Andrews statistics, Figure 20 is included. It represents the number of total installations as the ordinate and the total Kb/h traffic as the abscissa. All modes and their data rates are identified with two connected bubbles. The bubble with a T inside refers to total transmitted traffic for the appropriate mode. A bubble with an R inside stands for received volume.

From Figure 20 one finds that most installations belong to the 110 b/s, asynchronous, mode IIA category. However, most traffic is carried by 56.0 Kb/s, mode VI, followed closely by 19.2 Kb/s, mode VI. Based on AUTODIN II utilization alone, the ultimate mode would have to be selected from the above three.

Final Combined Selection

Only future developments, such as system growth, costs, implementation constraints, usage popularity, and actual importance of certain applications, may determine which mode has most utility. Likewise, technical vulnerability may depend on factors not foreseen today. However, certain relative ordering appears to be valid. The details of two such approaches have been given in (a) with respect to technical service vulnerability, and in (b) with respect to service utilization levels.

One may select three highest ranking candidates from (a) and (b). When this is done, only five finalists remain, because the 56.0 Kb/s, mode VI, is common to the two groups. The five finalists are:

Mode VI, 56.0 Kb/s
Mode IIA, 110 b/s
Mode IB, 56.0 Kb/s
Mode VI, 19.2 Kb/s
Mode VI, 1.2 Kb/s

Table 12. Andrews Utilization

Mode	Source Rate	Terminal Example	Number Install.	Total Traffic (KB/HR)	
				Transmitted	Received
IB	56.0 K	PROGTERMIN	2	70,234.4	193,348.8
	9.6 K	CDCUT 200	6	42,105.3	70,391.7
	4.8 K	IBM 2780	21	60,020.7	222,900.7
	2.4 K	BURTD 806	10	13,694.9	55,473.8
	1.2 K	DATA 100	3	855.3	5,467.7
	300	H700 KEYNET	3	729.8	2,048.8
	150	COPE 1200	1	82.4	412.2
IIA	2.4 K	SILENT 700	7	4,779.5	23,899.5
	1.2 K	KLP 300	19	10,873.4	54,377.0
	600	CRT/TTY	5	1,649.0	8,244.5
	300	HAZELT 2000	11	1,580.1	7,899.6
	150	COMPCOM 30	9	566.9	2,831.7
	110	TTY, ASR	90	3,083.5	15,704.1
IIAH	2.4 K	IBM 2741	4	4,319.1	21,595.4
	1.2 K	VIP 786 W	10	6,595.0	32,977.0
	300	UNISCOP 100	13	346.1	6,435.6
	110	BURTC 500	5	110.8	554.2
VI	56.0 K	HIS 115	12	851,156.3	538,626.6
	19.2 K	HIS 6000	20	460,991.9	359,392.4
	9.6 K	NOVA 800	7	75,654.4	57,961.9
	4.8 K	UNIVAC 494	10	90,345.4	81,661.6
	2.4 K	IBM 360/30	7	8,902.2	26,279.1
	1.2 K	BUR 3500	7	5,919.1	17,832.5

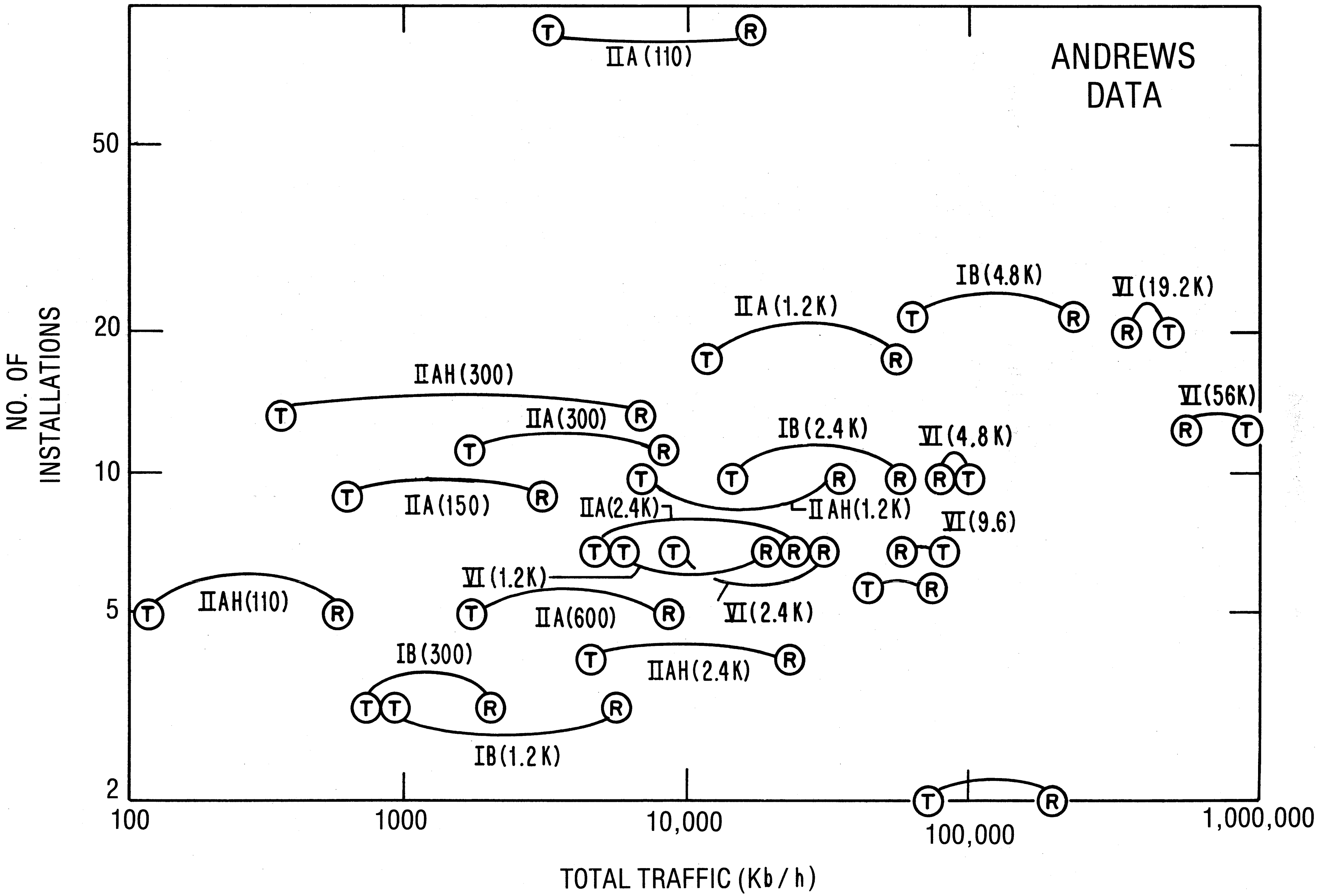


Figure 20. Andrews utilization.

Their relative ranking is summarized in Figure 21. On the basis of these results we have selected Mode VI, 56.0 Kb/s.

3.3 Selection of Analog Service Mode

Our concern here is to select one specific mode of operation for the analog service selected previously, namely the transfer of audio information (i.e., voice) between humans. We further assume for this mode of operation that the voice signals will be digitized at some point in the network. This is based on the ultimate objective of developing technical performance parameters which can be employed in the future design and specification of transmission facilities for the digital DCS. Also special attention should be given to secure voice conferencing which implies use of voice digitizing for encryption purposes.

The tabulation of candidate alternatives from which our selection was made is given in Table 13. We have included a conventional analog, clear voice mode as one alternative strictly for comparison purposes. All of the other alternatives involve some form of analog to digital (A/D) conversion process. Some of the candidates are distinguished by the location in the network where the A/D conversion occurs.

Alternatives 2 through 4 in Table 13 assume that the digitization process takes place either at the backbone switch or at the local access switch. The digitization technique employed may be PCM at 64 Kb/s, CVSD at 16 Kb/s, or linear predictive coding (LPC) at 2.4 Kb/s. The access loop to the backbone node is either analog transmission over a 4 kHz bandwidth or quasi analog transmission at the digitization rate. The backbone switch may use analog or digital technology depending on where the digital conversion occurs.

Alternatives 5 through 8 all assume that the digitization process is accomplished in the terminal station apparatus. In these alternatives the communication security (COMSEC) equipment may or may not provide encryption on an end-to-end basis. Subsets of each major alternative are distinguished by letters indicating analog (a) or digital (d) networks and clear or secure (s) transmission after digitization.

The rest of Table 13 is self-explanatory. It does not include any hybrid alternatives (e.g., digital switches with quasi analog transmission) or tandem mixes of different digitizing processes. These modifications would increase the alternative candidates to be considered to an excessive number. It was believed more appropriate to select one alternative from the basic types shown in the table and to consider others during later phases of the project.

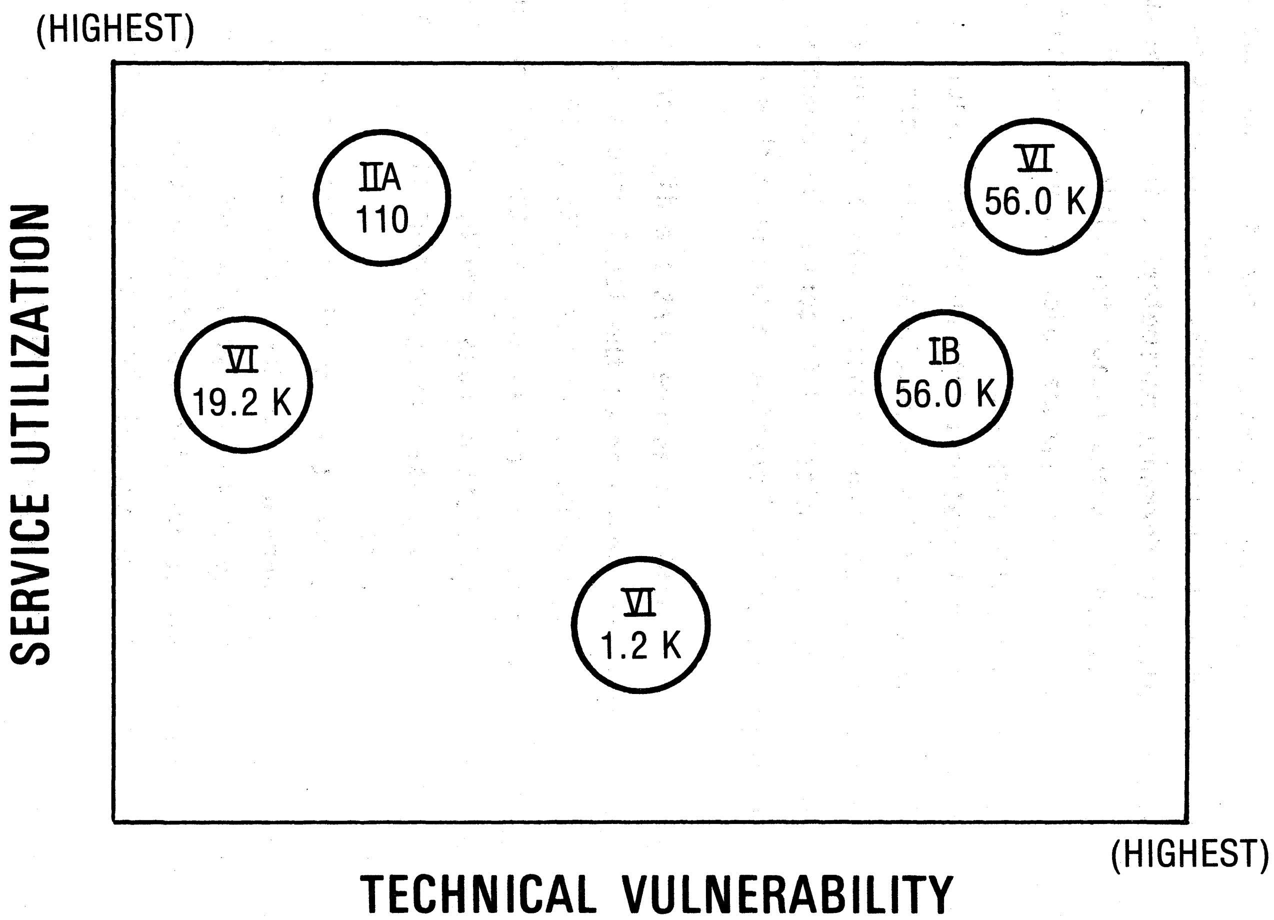


Figure 21. The five mode selection finalists.

Table 13. Alternative Configurations for Voice Service Modes

Alternative		Digitization/Encryption			Access	Loop	Switch Technology	Long Haul Transmission Backbone
Clear	Secure	A/D Location	Process	Rate	Trans	Bandwidth or Rate		
1a		None			Analog	4 kHz	Analog	Analog
2a	2as	Backbone	PCM	64 Kb/s	Analog	4 kHz	Analog	Quasi-Analog
2d	2ds	Switch	PCM	64 Kb/s	Analog	64 Kb/s	Digital	Digital
3a	3as	Backbone Switch	CVSD	16 Kb/s	Analog	4 kHz	Analog	Quasi-Analog
3d	3ds	Access Switch	CVSD	16 Kb/s	Analog	16 Kb/s	Digital	Digital
4a	4as	Backbone Switch	LPC	2.4 Kb/s	Analog	4 kHz	Analog	Quasi-Analog
4d	4ds	Access Switch	LPC	2.4 Kb/s	Analog	2.4 Kb/s	Digital	Digital
5a	5as	Station Apparatus	PCM	64 Kb/s	Quasi-Analog	Wideband	Analog	Quasi-Analog
5d	5ds	Station Apparatus	PCM	64 Kb/s	Digital	64 Kb/s	Digital	Digital
6a	6as	Station Apparatus	CVSD	16 Kb/s	Quasi-Analog	4 kHz	Analog	Quasi-Analog
6d	6ds	Station Apparatus	CVSD	16 Kb/s	Digital	16 Kb/s	Digital	Digital
7a	7as	Station Apparatus	LPC	4.8 Kb/s	Quasi-Analog	4 kHz	Analog	Quasi-Analog
7d	7ds	Station Apparatus	LPC	4.8 Kb/s	Digital	4.8 Kb/s	Digital	Digital
8a	8as	Station Apparatus	Vocoder	<2.4 Kb/s	Quasi-Analog	<4 kHz	Analog	Quasi-Analog
8d	8ds	Station Apparatus	Vocoder	<2.4 Kb/s	Digital	<2.4 Kb/s	Digital	Digital

Table 14 repeats the listing of alternatives from Table 13 and rates each criterion on the basis of five selection criteria discussed previously in Section 3.1 with the most weight given to vulnerability to transmission impairments and anticipated demand. Each alternative in Table 14 was considered under each selection criterion and subjectively rated as high, medium, and low. Under this procedure the preferred choice is the alternative with the highest rating. Alternative 6 in its various forms is the candidate finally selected. Alternative 6 is expected to have a high future demand for end-to-end secure voice transactions. The key to the use of 16 Kb/s CVSD for digital voice in the near future DCS is the availability of a reliable 16 Kb/s modem for transmission over narrowband 3 kHz to 4 kHz voice channels. The modem must provide satisfactory bit error rates on appropriate DCS channels normally used for AUTOVON. The USAF Rome Air Development Center (RADC) have supported the development and testing of such a modem over unconditioned 4 kHz analog channels. See McRae et al., (1976) and Perkins and McRae (1978).

Since CVSD terminals operate over four-wire full duplex circuits, two narrowband voice channels are needed. It is possible to operate CVSD in a conferencing mode. A conference bridge operation employs a typical tandem connection for 3-way conferencing. If one terminal is designated as the hub H, and the two conference terminals as X and Y, the voice paths H-X and H-Y are encoded and decoded only once. However the voice path X-Y requires two separate CVSD encoder/decoder operations. This tandem connection suffers in quality at the H-X or H-Y connection, because delta modulation, including CVSD, does not tandem as well as PCM.

Transmission at 16 Kb/s is accomplished at a symbol rate of 2.667 Kb/s, thus requiring 6 bits per symbol. The transmitted format is a suppressed carrier quadrature amplitude modulation (QAM) with 64 possible symbols. These 64 symbols are derived from all combinations of 16 phase and 4 amplitude levels. An adaptive equalizer permits operation over a number of line conditions by compensating for amplitude and envelope delay characteristics. The compensation achieved depends on the line characteristics. A modest initialization period is required by the equalizer. Transmission requirements to support such a modem are expected to be fairly stringent. Therefore, the technical channel criteria used for the 16 Kb/s modem should provide satisfactory service for other modes of narrowband voice operation. This is one of the basic reasons for selecting the high speed mode of operation over narrow bandwidth channels. Ultimately the voice quality measures developed would be applied to the 16 Kb/s CVSD voice digitization process over a

Table 14. Basis for Alternative Selection

Basis for Selecting / Alternative	Anticipated Demand	Military Operations Impact	Future Implementing Possibility		Influence on Trans Rqmts		Data Base Availability	Comments
			Near Term	Long Term	BW	Error Rate		
1a	High	Medium	Low	Low	Low	N/A	High	Conventional Use AUTOVON I
2a, 2as	Medium	Medium	Med	Med	High	Low	High	Some in Use (DEB)
2d, 2ds	High	Medium	High	Med	High	Low	High	CONUS AUTOVON (T-Carrier)
3a, 3as	Low	Low	Low	Low	Med	Med	Med	Limited Use (Mostly Tactical)
3d, 3ds	Low	Low	Low	Med	Med	Med	Low	Limited Use (Tactical)
4a, 4as	Low	Low	Low	Low	Low	High	Low	Relatively New
4d, 4ds	Low	Low	Med	Med	Low	High	Low	Relatively New
5a, 5as	Low	Medium	Low	Low	HIgh	Low	Low	Limited Use
5d, 5ds	Low	Medium	Low	Low	High	Low	Low	Limited Use
6a, 6as	High	High	Med	High	Med	Med	Med	Some Tactical & SVIP
6d, 6ds	High	High	Med	High	Med	Med	Low	Some Tactical & SVIP
7a, 6as	Medium	High	Low	Med	Low	High	Low	New Technology
7d, 7ds	Medium	High	Med	High	Low	High	Low	New Technology
8a, 8as	Low	Low	Low		Low	High	Low	Special Application (Synthetic Quality)
8d, 8ds	Low	Low	Low		Low	High	Low	Special Application (Synthetic Quality)

16 Kb/s voiceband transmission facility. In this connection, it should be noted that proponents of CVSD maintain that 48 Kb/s CVSD is equal in quality to 64 Kb/s PCM. Voice encoded at 64 Kb/s PCM is generally regarded as high or toll quality, particularly for tandem connections. Critics claim that CVSD does not operate in tandem as well as PCM, as noted earlier. Military users have observed excellent quality for 32 Kb/s CVSD. At 16 Kb/s CVSD loses tonal quality and degrades other secondary quality features, but retains intelligibility and speaker recognition, the primary acceptance criteria. For these reasons, CVSD has been selected for use in tactical systems such as TRI-TAC. When combined with the 16 Kb/s modem, CVSD becomes highly desirable since it can operate over existing voice frequency channels. Also, of course, the digital voice can be easily encrypted for secure transmission on an end-to-end basis.

4. SELECTION OF DIGITAL SERVICE PARAMETERS

4.1 Overview of Performance Parameters

Telecommunications systems provide a variety of digital services to a growing group of users or subscribers. The performance of these digital services has been described in many ways (AT&T, 1970; ANSI, 1971; Mahoney, et al., 1975; Frank, et al., 1976; Audin, 1978; Seitz and McManamon, 1978; Kleinrock, 1978; Grubb and Cotton, 1978; ANSI, 1980). The reason for this existing dichotomy has to do largely with individuals and organizations who have different responsibilities, different requirements, and different views of the service. As noted in Section 2, examples of divergent viewpoints may be found between system designers, system operators, traffic managers, and different priority end users.

The divergence of viewpoints has given rise to division of the parameter space into at least two groups: the system oriented and the user oriented parameters. The system or system oriented parameters are also called technical or engineering parameters, in this report and elsewhere. In the discussion of digital service, we prefer the adjective "system". This will avoid confusion in Sections 4 and 5, where the typical user will turn out to be a technical device, such as a high speed automatic data processing (ADP) installation, a host computer, or an access software to another network. The service received by such users will be, unfortunately rather ambiguously, described in both technical and engineering terms. One should also note that user oriented and service oriented parameters denote the same family of descriptors, at least in this text. Whether a certain parameter, such as quality of service, belongs to the system or user group depends on the definition. Generally, however, systems oriented parameters refer to the performance of either the entire system, its subsystems, functions, facilities, its hardware, software, or any combinations of the above. The user oriented or service oriented parameters, on the other hand, describe the service as perceived by the end users. Later in this section, different end user definitions will be discussed. Depending on end user-system interface, for instance, a service parameter such as access delay could have different interpretations, as well as different numerical values.

Consider, for example, link outage probabilities in a microwave network. By themselves, they appear to be system parameters. Because, given sufficient reroute topology and control capability, the majority of outages may not be experienced by the end users. On the other hand, nonredundant networks may translate link outages into clear-cut service outages. The user thus experiences the effects of the entire system, including links, nodes, topologies and controls, as a whole.

Since the inception of digital telecommunication services several decades ago, the relative advantages and disadvantages of offered services have been discussed by many. With or without regard to parametric classification, a great many descriptors have been used to qualify and to quantify various services. At times different words have been used to refer to the same feature. The same words have also been used for different features.

The multiplicity of words is by no means the entire problem. Interpretations can and do vary. The same parameter terminology can be defined to pertain strictly to service here and now, but to the system elsewhere. To arrive at such multidefinitions, it suffices to displace the user-system interface in space and time.

The majority of existing performance descriptors tend to combine in some intrinsic way the aspects of both service and system. Consider Table 15. It lists a total of 66 parameters. The listing is done in alphabetical order. Hence, no significance is to be assigned to a parameter appearing in first, second, or last position. Some parameters may be unique or distinct. Several others may refer to the same thing - fully or in part. Furthermore, while some parameters pertain strictly to the service, the system, or the operator performance, the majority of common parameters appears to combine somehow the orientation of user, system, or operator.

This issue is addressed in the three right-hand columns of Table 15. These columns, called Parameter Orientation, include a check (x) for user orientation, system orientation, or the previously mentioned operator orientation. The accuracy

Table 15. A Listing of Assorted Performance Parameters

#	Performance Parameter	Parameter Orientation		
		User	System	Operator
1.	Accuracy	X		
2.	Adapatibility	X	X	X
3.	Availability	X		
4.	BER end-to-end	X		
5.	BER subsystem (link, etc.)		X	
6.	Bit count integrity	X	X	
7.	Block (char., etc.) count integrity	X	X	
8.	Blocking probability (access)	X		
9.	Blocking probability (nodes)		X	
10.	Communications error	X	X	X
11.	Compatibility (facility, function)	X	X	X
12.	Congestion duration	X	X	X
13.	Congestion frequency	X	X	X
14.	Cross-connect constraints	X	X	
15.	Delay in backbone delivery		X	
16.	Delay in end-to-end delivery	X		
17.	Delay in system access	X		
18.	Destructibility (hardware, software)	X	X	X
19.	Duplication probability of blocks	X	X	
20.	Echo options	X		X
21.	Efficiency	X	X	X
22.	Error-control options	X	X	
23.	Error-free block ratio	X	X	
24.	Error rate for output blocks	X		
25.	Error rate for subsystem blocks		X	
26.	Flexibility	X	X	X
27.	Format interfacability		X	X
28.	Grade of service (GOS)	X	X	X
29.	Maintainability		X	X
30.	Misdelivery probability of blocks	X		
31.	Missing block ratio	X		
32.	Operability			X
33.	Outages at service access	X		

Table 15. (Continued)

#	Performance Parameter	Parameter Orientation		
		User	System	Operator
34.	Outages of subsystems		X	X
35.	Performance end-to-end	X		
36.	Performance of subsystems		X	
37.	Precedence and preemption capability	X		
38.	Priority level management	X	X	X
39.	Privacy	X	X	X
40.	Quality of service (QOS)	X	X	X
41.	Rate of effective bit transfer	X		
42.	Rate throughput for blocks	X		
43.	Reliability	X	X	X
44.	Restoration of service options	X	X	X
45.	Robustness	X	X	X
46.	Security end-to-end	X	X	X
47.	Security of subsystems		X	
48.	Service classes	X		X
49.	Service features	X		
50.	Service quality	X		
51.	Subsequent segment acceptance	X		
52.	Survivability	X	X	X
53.	Systemwide security	X	X	X
54.	Text code transparency	X		
55.	Throughput rate constancy	X	X	
56.	Time between bit count losses		X	
57.	Time to service restoral (MTSR)	X		X
58.	Tracing options for lost packets		X	X
59.	Transmission activity security		X	
60.	Transmission signaling security		X	
61.	Transparency	X	X	X
62.	Unavailability probability (≥X s)	X	X	X
63.	Upgradeability of user services	X		
64.	Vulnerability to subsystem failures		X	X
65.	Vulnerability to operator errors			X
66.	Vulnerability to user errors	X		X

parameter, for example, is checked as user oriented. The system and the operator must, by all means, perform statisfactorily to provide the accuracy level in question. But it is the user who is primarily concerned with the accuracy of his information transaction.

The adaptability parameter poses a different problem. If the service cannot adapt to different traffic demands placed by one or more users, it may be restrictive to that user and others. At the same time, there may be many technical issues involved in system adaptation to new services and new technologies. Finally, the operator may or may not play a key role in the entire adaptation process. In Table 15, adaptability is shown to be associated with all three orientations.

One can proceed in a qualitative way from the top to the bottom of Table 15. In most cases, the distinction may not be clear, but some decision appears warranted. The x's in the table reflect the subjective judgement of the authors. It is based on one's perception of what constitutes a user, a system, or an operator; as well as on the often informal definition of what the words should mean for this or that parameter.

Since all past and present military, government, civilian, and commercial communication systems have used at least some of the 66 descriptors of Table 15 applied, it seems worthwile to scrutinize at least some of them in more detail. Let us consider the binary or bit error rate (BER) and the quality of service (QOS).

Binary error rate or probability is perhaps the most researched and measured digital service parameter. It has been derived for individual links, such as transmission channels of all kinds imaginable, and subject to innumerable degradations and distortions. These BER studies have extended to feedback channels and to network topologies of growing complexity. Different combinations of links, nodes, and control processes have been included in the studies. A simple tandem arrangement is illustrated in Figure 22.

Errors can originate, or at least be initially observed, almost everywhere. To claim that a particular error has occured in a user terminal, and neither on a local loop nor in the switch (e.g., PSN), one needs to resolve the matter of interfaces. Where does the terminal end and the loop start? Where is the interface between the local loop and the switch? Likewise, to discriminate between user generated errors and those produced by the system, one needs to identify a unique boundary, called the user-system interface.

The most familiar standard interface is the so-called data terminal equipment (DTE) and data circuit-terminating equipment (DCE) interface. Its physical

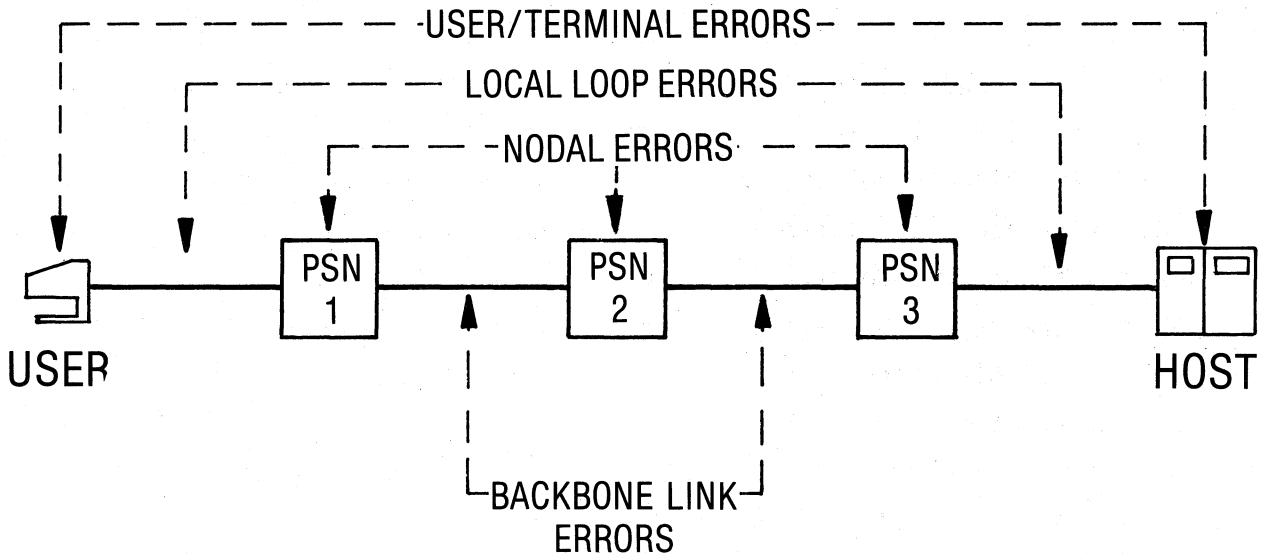


Figure 22. Potential physical sources of bit errors in a tandem message path.

requirements are specified in the industry standard X.21 (ANSI, 1980). In a more general setting, such as for packet modes on public data networks, the DTE/DCE interface falls in the domain of standard X.25. As shown in Figure 23, it separates, at least conceptually, the terminal and communication equipment. Several questions remain. What if the same processor acts at one time as terminal equipment and at other times as communications equipment? Furthermore, what about the human user who leases the entire service, including the DTE, from a communications carrier? As we shall see later in the discussion of the Interim Federal Standard 1033, such questions can be resolved by a standard set of interface definitions.

Three potential user-system interfaces are shown in Figure 24. The BER could be measured at all three interfaces and, unless one is lucky, they could differ at the three places. One could venture that the interface nearest to the human operator, i.e., the outermost or the "FS-1033" interface, would have the highest BER of the three. This may be true in installations that incorporate no error control. However, in systems with error control, such as with a cyclic redundancy check (CRC) plus an automatic repeat request (ARQ), the situation could be reversed. The "line side" interface in Figure 24 is shown to be the same as the DTE/DCE interface (see Figure 23). In the host computer packet applications, it may be equivalent to a demarcation line between the transmission control program (TCP) and the segment interface protocol (SIP), both of which are shown to be parts of the single channel control unit (SCCU). If the DTE can be easily distinquished from the user terminal, another interface - called "optional" in Figure 24 - may also be considered to separate the host computer from the host specific interface (HSI) of the SCCU.

The user-system interface is needed to distinquish between the end-to-end, user-oriented BER and the subsystem (viz., link, node) BER's in Table 15. The end-to-end BER (see #4 in the table) is said to be user oriented. It is measured at the receiving end of the data path, as the data pass through the system-user interface, and arrive at the end user. On the other hand, the subsystem BER (item #5) is called system oriented. It occurs somewhere in the system, usually a considerable distance from the user-system interface. To discern where exactly the subsystem errors arise within the tandem structure of a network (see Figure 22), one may establish appropriate interfaces, sometimes of the DTE/DCE type, between nodes and links on the message transfer path.

Another difficulty arises when two similar DCE's, or their equivalents, interact at a gateway of two networks with no apparent DTE separating the two. In such cases a common error control arrangement may penetrate the two networks. On the

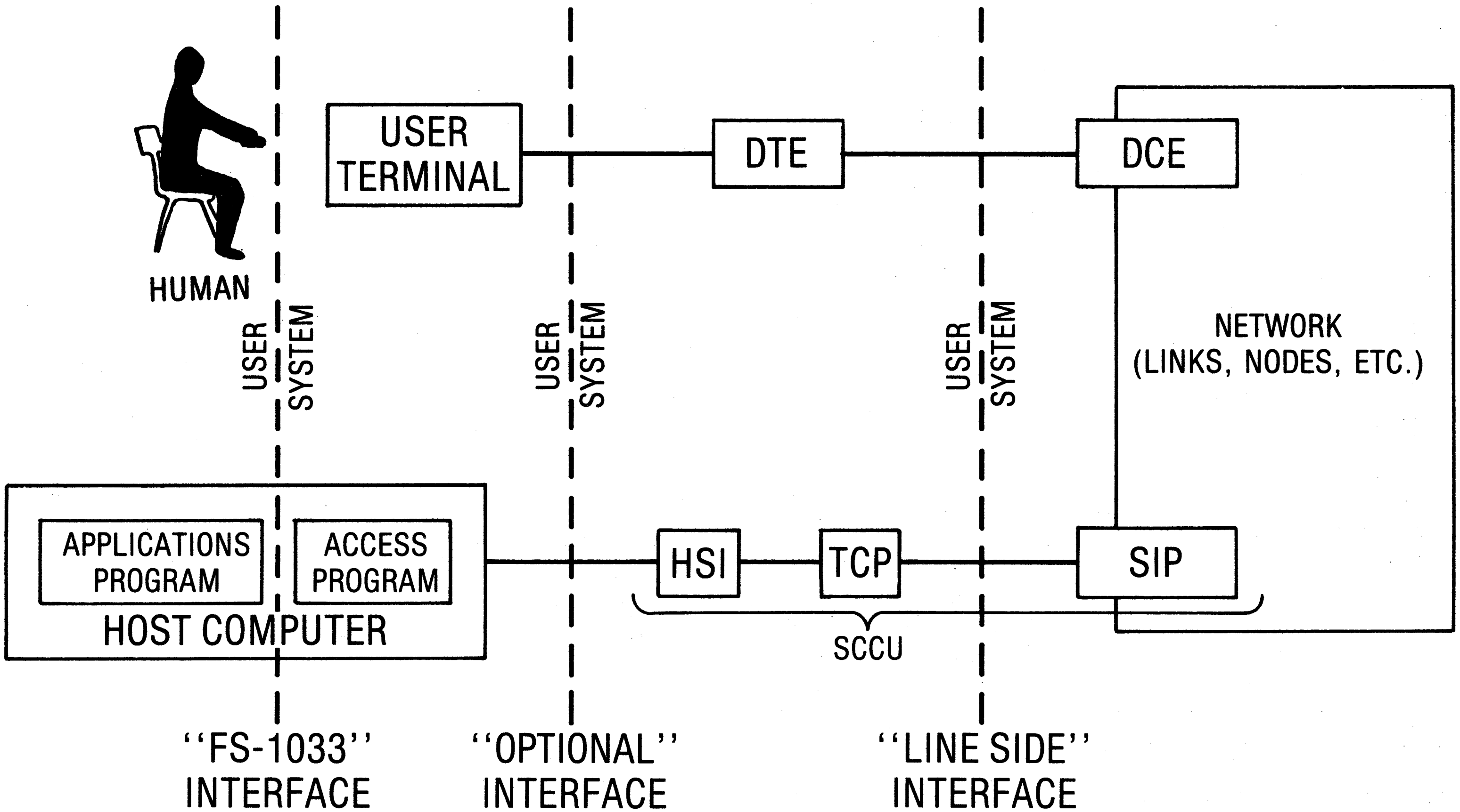


Figure 24. Three potential user/system interface definitions.

physical gateway channel then the BER could exceed the end-to-end BER's delivered to either network. The subsystem BER's are tied to the system structure. Ultimately, that structure determines the end-to-end or end user BER. In this study, emphasis is placed on this end-to-end or user BER.

The term "quality of service" (QOS) refers to more than the probability of blocked access called "grade of service" (GOS). Given some fixed waiting time, T, QOS stands for the probability that access is delayed more than T seconds. Thus, in common circuit switching use, GOS is viewed as QOS with the particular value T = 0. In packet switched networks, access delays occur when the originating PSN refuses to accept an incoming packet. This may happen for several reasons, such as incorrect format, identification, address fields, and so on. However, when the packet appears to be valid and the blockage is caused by full buffers at a congested PSN, then one can count the access delay against QOS. There are always two ways to look at an inadequate QOS. Since larger node and link capacities would normally ameliorate blocking and delays, the system can be viewed as underdesigned. Poor QOS would then be a fault of the system. On the other hand, given a fixed network capacity, there may be no noticable blocking if the offered traffic stays under some design bounds. Blocking and delays can arise when traffic volumes exceed these bounds. Now QOS depicts unpredictably high message activity by the subscribers. It seems that an acceptable view must be some compromise between the above extremes. The interaction of user demands with system capacities cause congestions. The resultant service degradation is usually quite apparent to both system users and operators. In Table 15, item #40, QOS, is indicated to have joint user, system, and operator orientation.

It is apparent that every one of the 66 parameters in Table 15 contains a story similar to the examples discussed above. Each parameter must be defined subject to its selected interfaces. After that, the parametric orientation with respect to user, system, and/or operator must be researched and assessed. The end result, whether qualitative or quantitative, would be long and undesirable. The parameter set is (a) too big, and (b) a perplexing mix of user, system, and operator factors. Since one is largely interested in user oriented parameters here, one seeks a reduction of Table 15 that suits the needs of the military digital service user.

4.2 Interim Federal Standard 1033

In this section we review the recently proposed standard for depicting digital communications performance, the Interim Federal Standard 1033, or FS-1033 for short.

In later sections we shall enlarge on two issues of FS-1033, namely on the separation of user and system parameters, as well as on the relationships between the parameters proposed in FS-1033 and those of immediate concern to AUTODIN II.

FS-1033 is currently an interim standard issued by General Services Administration (GSA). Its use is optional for all Federal agencies. The purpose of the standard is to improve Federal Government procurement and services by facilitating interoperability between telecommunication systems and information processing systems of the United States Government. To satisfy this broad mandate, FS-1033 proposes a common, user-oriented performance parameter set. The set is called uniform, to mean that it applies uniformly and without exception to all digital services.

The Interim Federal Standard 1033 has been studied by other standards organizations. It is the basis for developing a Proposed ANSI Standard, User-Oriented Data Communication Performance Parameters, X3S35/125.

The standard defines 26 digital communication performance parameters. Each parameter is to address a particular facet of telecommunication system performance, eventually in quantitative terms, and strictly from the viewpoint of the end user. The parameters are not restricted, either in definition or in application, to particular digital system classes.

The standard does not specify or suggest numerical values for the 26 parameters. Requirements and methods for measuring actual telecommunication system performance in terms of these parameters are currently being developed by NTIA/ITS and NBS. That work will lead to a separate Proposed Federal Standard, FS-1043.

It should be emphasized that FS-1033 is intended for all Federal departments and agencies. When ultimately approved on a permanent (non-interim) basis, the standard will be required for use by the Federal agencies in specifying the end-to-end performance required of all digital telecommunication systems and services for which planning and design begin more than one year after the effective date of the standard.

The standard is not intended to eliminate or to restrict the use of additional parameters. However, such additional parameters should not be used in lieu of any of the 26 parameters defined in the FS-1033.

The 26 service performance parameters of FS-1033 are listed in Table 16. The detailed definitions of all 26 parameters are given in the actual Interim Federal Standard 1033, a description of which is enclosed in this Report as Appendix B.

A summary introduction of the parameters is given next, followed by definition of the key terms. The 26 parameters of FS-1033 are divided into three parts: A, B,

Table 16. Service Performance Parameters According to Interim Federal Standard 1033

	Part A - Primary Parameters		
1.	Access Time	________	Seconds
2.	Incorrect Access Probability	________	*
3.	Access Denial Probability	________	*
4.	Bit Transfer Time	________	Seconds
5.	Bit Error Probability	________	*
6.	Bit Misdelivery Probability	________	*
7.	Bit Loss Probability	________	*
8.	Extra Bit Probability	________	*
9.	Block Transfer Time	________	Seconds
10.	Block Error Probability	________	*
11.	Block Misdelivery Probability	________	*
12.	Block Loss Probability	________	*
13.	Extra Block Probability	________	*
14.	Bit Transfer Rate	________	Bits/Second
15.	Block Transfer Rate	________	Blocks/Second
16.	Bit Rate Efficiency	________	%
17.	Block Rate Efficiency	________	%
18.	Disengagement Time	________	Seconds
19.	Disengagement Denial Probability	________	*
	Part B - Secondary Parameters		
20.	Service Time Between Outages	________	Hours
21.	Outage Duration	________	Hours
22.	Outage Probability	________	*
	Part C - Ancillary Parameters		
23.	User Access Time Fraction	________	*
24.	User Block Transfer Time Fraction	________	*
25.	User Message Transfer Time Fraction	________	*
26.	User Disengagement Time Fraction	________	*

*Note: The probabilities and user performance time fractions are dimensionless numbers between zero and one.

and C. As noted in Table 16, Part A is comprised of 19 primary parameters. Part B consists of 3 secondary parameters. Finally, Part C contains 4 so called ancillary parameters. The ancillary parameters provide a quantitative means for expressing the influence of user performance or nonperformance (e.g., user caused delays) on the primary parameter values. As noted earlier, the separation of system and user roles is not always easy or even feasible. There is also the obvious problem in employing user-dependent parameters to specify required system performance: the system operator normally has no control over user performance, and hence cannot ensure that user-dependent parameter values will be met.

More specifically, the FS-1033 ancillary parameters deal with user delays. The transaction functions (Seitz and Bodson, 1980) are structured in such a way that they involve a sequence of interactions between the system and the users. The overall performance times thus depend jointly on both system and user delays. The FS-1033 approach divides the resultant performance times into system and user fractions. The standard defines the average user fraction as an ancillary parameter which modifies the associated primary performance parameter. Rather fortunately, the digital 56.0 Kb/s mode selected for this study (see Section 3.2) appears not to suffer any significant effects from the four ancillary user fractions.

For the purpose of this study, Parts A and B are deemed far more important than Part C. Therefore, Part C (i.e., the 4 ancillary parameters) are not emphasized here. The secondary parameters deal with outages. They appear sufficiently important to most military communication services to retain a key role in AUTODIN II digital performance specification.

There are 19 parameters of the primary class. To facilitate their discussion, FS-1033 divides them further into three functional phases. The phases are called access, information (e.g., bit, block, message) transfer, and disengagement. This division into phases is illustrated in Table 17. Note that information transfer is the domain of 14 parameters, or more than half of the total. However, one may also note that half of the 14 pertain to bit transfer, while the other half pertain to block transfer. Blocks are defined as contiguous aggregates of bits that are treated as units. Examples of blocks may be characters, words, packets, segments, text fields, and so on. Often the number of bits in a block is a constant.

The bit and block transfers may have distinct performance categories or criteria. In FS-1033, these criteria are called efficiency (or speed), accuracy, and reliability. As one expects, such criteria are sufficiently general concepts to apply almost everywhere. In Table 18 one finds the FS-1033 interpretation of the

Table 17. The Access, Information Transfer and Disengagement Phases for the Primary Parameters

Access Phase

1. Access Time
2. Incorrect Access Probability
3. Access Denial Probability

Information Transfer Phase

4. Bit Transfer Time
5. Bit Error Probability
6. Bit Misdelivery Probability
7. Bit Loss Probability
8. Extra Bit Probability

    9. Block Transfer Time
    10. Block Error Probability
    11. Block Misdelivery Probability
    12. Block Loss Probability
    13. Extra Block Probability

14. Bit Transfer Rate

    15. Block Transfer Rate

16. Bit Rate Efficiency

    17. Block Rate Efficiency

Disengagement Phase

18. Disengagement Time
19. Disengagement Denial Probability

Table 18. The Three Performance Criteria

Function	Performance Criterion		
	Efficiency or Speed	Accuracy	Reliability
Access	1. Access Time	2. Incorrect Access Probability	3. Access Denial Probability
Bit Transfer	4. Bit Transfer Time	5. Bit Error Probability 6. Bit Misdelivery Probability 8. Extra Bit Probability	7. Bit Loss Probability
Block Transfer	9. Block Transfer Time	10. Block Error Probability 11. Block Misdelivery Probability 13. Extra Block Probability	12. Block Loss Probability
Message Transfer	14. Bit Transfer Rate 15. Block Transfer Rate 16. Bit Rate Efficiency 17. Block Rate Efficiency	22. Outage Probability (Secondary)	
Disengagement	18. Disengagement Time	19. Disengagement Denial Probability	

Service Continuation	20. Service Time Between Outages (Secondary)
Service Restoral	21. Outage Duration (Secondary)

criteria as further specified for the 22 primary and secondary parameters. The three secondary are region numbers 20, 21, and 22 in the table. Note, that along the left margin the communications phases are expanded and modified. In FS-1033, these phases or communications activites are not the previous three (i.e., access, message transfer and disengagement), but rather a composite of seven so-called functions. Each function, as depicted in Table 18, consists of parametric efficiency, accuracy, and reliability components. For instance, the access function has Access Time (a primary parameter) as its efficiency parameter, Incorrect Access Probability (a primary parameter) as its accuracy parameter, and Access Denial Probability (also a primary parameter) as its reliability parameter. For the message transfer function, which is defined to differ from both bit and block transfer functions, the efficiency criterion is claimed to contain four primary parameters: Bit Transfer Rate, Bit Rate Efficiency, Block Transfer Rate, and Block Rate Efficiency. Furthermore, the accuracy and reliability criteria of the message transfer function are jointly covered by the secondary parameter, Outage Probability.

Two unique communication functions in Table 18 are service continuation and service restoral. They each possess a single secondary parameter, namely Service Time Between Outages and Outage Duration, which make no apparent discrimination between the criteria of efficiency, accuracy, and reliability.

Being new and quantitative, the FS-1033 parameters have been dilligently defined. The definitions are based on the above precepts of functions and criteria. They also rely on the existence of uniquely defined user/system interfaces. More on the user-system separation will be said in the next section. As noted earlier, see Figure 24, the FS-1033 places the user-system interface closer to the user than other definitions.

The key point is that, in its simplest human operator case, the user-system interface coincides with the finger-eye-ear interface of the human with the terminal console. There are, of course, many more complicated interface cases. They include separation of interactive end user actions from system actions; information separation of user data from system overhead (signaling and control); the functional definition of the end user in the instances of unattended devices; joint applications and communications programs; intermediate devices or processes, that may be acting as gateways between different networks; and the distinction between local data processing and so-called telecommunications access methods.

To use the FS-1033, a four-part overview may be helpful. The four parts are:

(1) The user-system model

It defines the end users or sources, the data communications system, and the interfaces between users and the system.

(2) The activity phases

Each communications action may be viewed as a performance trial. It has its time window, a start and an end. Called functions in the FS-1033, the main parts of the time window are access, information transfer, and disengagement phases.

(3) Performance criteria

Each trial has an outcome. The goodness quality of that outcome is judged first rather broadly by three performance criteria: efficiency -- to describe successful performance, accuracy -- for incorrect performance, and reliability -- for nonperformance.

(4) Performance parameters

For each of the three criteria, during every phase (function) of the activity window, the trial outcome is to be described by specific performance numbers. These numbers are the 26 parameters of FS-1033: 19 primary, 3 secondary, and 4 ancillary. The numbers represent time (seconds), rates (bits or blocks/second), efficiencies (%), or probabilities (dimensionless), as shown above in Table 16.

It is also advisable to employ a single correct list of terms and definitions. To this end, the Interim Federal Standard 1033 has an extensive list of definitions. The proposed ANSI Standard X3S35/125 has a slightly modified listing of definitions and explanations throughout its text and appendices. Finally, a less formal outline of FS-1033 and its applications is given by Seitz and Bodson (1980) or Appendix B.

4.3 The User and System Parameters

The separation of user oriented parameters from system oriented parameters is the main objective of this study, as well as that of FS-1033. The demarcation is achieved with the aid of the user-system interface. For the purposes of this discussion, the interface is placed as illustrated in Figure 25. The FS-1033 model defines the end user of AUTODIN II data communications system or service either as a human terminal operator, a communications applications program, a gateway access program in another network (such as AUTODIN I), or an unattended device (such as a card puncher or message recorder).

As noted in Section 4.1, another factor between end user and the system proper may be the system operator or the communications manager. We shall assume the following. If the operator is also a user, he or she shall be called a user. If the operator is not an end user, he or she shall be identified, at least functionally,

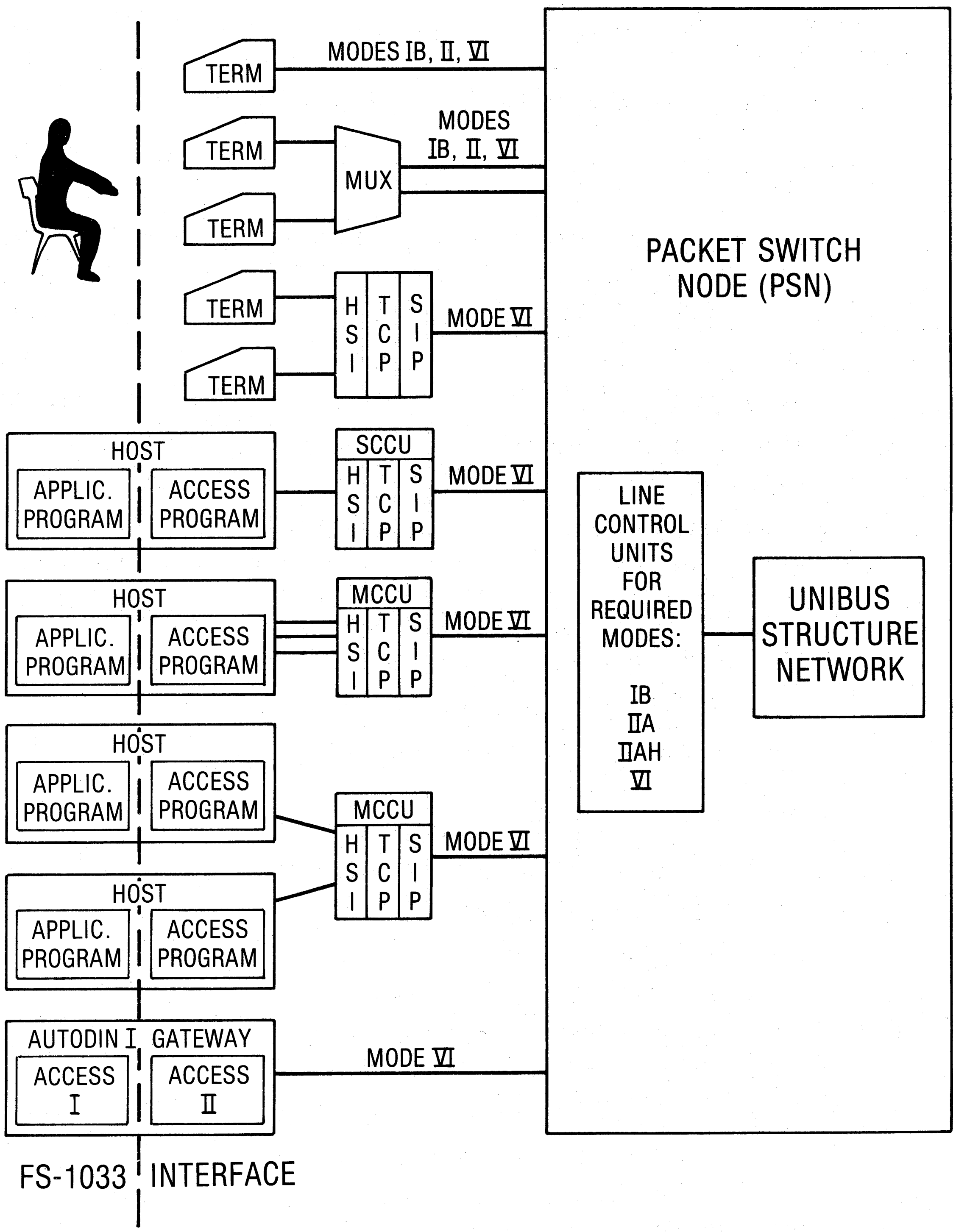


Figure 25. The FS-1033 user/system interface at an AUTODIN II node.

as part of the system. From now on, accordingly, one must worry only about the user vs. the system.

In Table 15, more than 60 assorted performance parameters are listed. To render this list more manageable, one can use the following three-step procedure:

(i) Retain only those parameters that emphasize service received by the end user through the user-system interface.

(ii) Merge related or overlapping parameters.

(iii) Delete all vague, qualitative, marginally relevant, or insignificant parameters.

For example, the very first item in Table 15 is accuracy. It seems to encompass many other parameters, such as BER end-to-end (#4), BER subsystem (#5), bit count integrity (#6), block count integrity (#7), communications error (#10), duplication probability of blocks (#19), echo options (#20), error control options (#22), error-free block ratio (#23), etc. According to (iii), accuracy is judged to be so vague and qualitative, as to be nearly meaningless for quantitative use. It is deleted from further consideration. Likewise, the second parameter in Table 15, adaptability, is deleted because of reasons (i) and (iii). Clearly, one may have to adapt to new subsystems, new technologies, new types of message traffic, new service features, as well as new user requirements. Adaptability emphasis does not seem to be on the user orientation.

The end result of this deletion and merger process is a shorter list that stresses the service received by the end user. The abbreviated listing is given in Table 19. It has 16 elements.

The 16 parameters are called "common" to emphasize their frequent occurrence in various civilian and military situations. These parameters are user oriented, to be observed at the end user side of the user-system interface.

4.4 Relationship Between FS-1033 and AUTODIN II Parameters

The goal of this section is to compare FS-1033 and AUTODIN II digital service parameters. The user oriented FS-1033 parameter set will be comprised of the primary and secondary parameters. See previous Section 4.2 for their definitions. The AUTODIN II service parameters are much less defined, to our knowledge. The system performance requirements in the AUTODIN II design plans and its service specification combine in various ways the user, the system, and the operator aspects. The 66 parameter listing of Table 15 seems to contain most, if not all, of AUTODIN II performance parameters. Clearly, this list is too long to be

Table 19. Common User Oriented Performance Descriptors

#	Descriptors (listed alphabetically)
1.	Availability of service
2.	BER end-to-end
3.	Bit count integrity
4.	Block count integrity
5.	Block error rate end-to-end
6.	Blocking/delay probability (GOS/QOS)
7.	Mean delay at access
8.	Mean delay end-to-end
9.	Misdelivery probability
10.	Next segment acceptance probability
11.	Outage probability
12.	Throughput rate for bits
13.	Throughput rate for blocks
14.	Time between loss of bit/block counts
15.	Time between outages (MTBO)
16.	Time to service restoral (MTSR)

useful. Furthermore, there are questions about the parameter definitions, their dependence on elements beyond user control (e.g., system, operators), mutual overlap, AUTODIN II applicability, and realistic usefulness in practice.

To simplify matters, and to keep the user orientation in focus, this section will assume the 16 so-called "common performance parameters" of Table 19 as representative general descriptors of military digital service. They are particularly applicable to the AUTODIN II service. This is somewhat restrictive, as there certainly are more user-related parameters being used than the 16 listed. If one finds the abbreviated list inadequate, one can always add additional parameters or use any longer list, like Table 15.

One next compares the primary and secondary performance parameters of FS-1033 (see Tables 16 and 18) with the abbreviated and condensed 16 parameter list (Table 19). As a starting point, consider the coincidence matrix shown in Table 20. This is essentially a 22 x 16 array, having as rows the FS-1033 parameters and as columns the common parameters of Table 19. For lack of space, the parameter names are not reproduced along margins. Rather, the same parameter numbers are listed as they appear in the respective tables. For instance, in FS-1033 Table 16, #5 is Bit Error Probability, and in the common parameter Table 19, #5 is Block error rate end-to-end. At the intersection of each row and column, the entry denotes the cross-correlation, coincidence of content, or the common technical meaning.

Three entries are used. Capital X stands for a full agreement or match for the two margin parameters. Capital P denotes partial agreement in at least one of the axes. A void entry means that there is insignificant cross-correlation or none at all.

The void areas are useful to identify incompleteness of the respective parameter sets, especially if they extend all the way across the table. To illustrate this issue, Table 21 accentuates the void rows and columns. It is easily seen from the rows that three FS-1033 parameters:

2. Incorrect Access Probability,
18. Disengagement Time,
19. Disengagement Denial Probability,

have no immediate counterpart in the common parameter set. This is deemed to be a shortcoming of Table 19, since these three parameters are quite pertinent to AUTODIN II performance characterization.

By looking at the vacant columns one recognizes that there are descriptors in Table 19, namely:

Table 20. Coincidence Matrix for the Common and FS-1033 Service Parameters

Primary and Secondary Parameters from FS-1033		Common User Parameters from Table 19: 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Access	1							X									
	2																
	3	P					P										
Information Transfer - Bits	4								P								
	5		X														
	6									P							
	7			P													
	8			P													
	14												P				
	16												P				
Information Transfer - Block	9								P								
	10					X											
	11									P							
	12				P												
	13				P												
	15													P			
	17													P			
Diseng.	18																
	19																
Outages	20															X	
	21	P															P
	22	P										P					

Table 21. Relative Incompleteness of the Two Parameter Sets

Common User Parameters from Table 19 (columns); Primary and Secondary Parameters from FS-1033 (rows)

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Access	1							X									
	2																
	3	P					P										
Information Transfer - Bits	4								P								
	5		X														
	6									P							
	7			P													
	8			P													
	14												P				
	16												P				
Information Transfer - Blocks	9								P								
	10					X											
	11									P							
	12				P												
	13				P												
	15													P			
	17													P			
Diseng.	18																
	19																
Outages	20															X	
	21	P															P
	22	P										P					

10. Next segment acceptance probability,
14. Time between loss of bit/block counts,

that appear to have no obvious counterpart in FS-1033. This may be a weakness of the standard, but not a substantial one.

Next segment acceptance or nonacceptance, if viewed as a new trial, could be counted in FS-1033 as part of the three parameters constituting the access phase. In particular, it appears to be closest to item 3, Access Denial Probability, in the new trial sense. If viewed as an ongoing trial, the failure to accept the next message segment would fall under FS-1033 parameter 22, Outage Probability.

Likewise, time between loss of bit/block counts is not that basic to the end user. One way or another, its effects are destined to materialize in one or more of the following degradations (in the FS-1033 sense):

6. Bit Misdelivery Probability,
7. Bit Loss Probability,
8. Extra Bit Probability

or

11. Block Misdelivery Probability,
12. Block Loss Probability,
13. Extra Block Probability.

One concludes from the vacant rows and columns that FS-1033 is more complete than Table 19.

Consider next the X entries in Table 20. There are four of them. From FS-1033:

1. Access Time,
5. Bit Error Probability,
10. Block Error Probability,
20. Service Time Between Outages,

appear to have a one-to-one mapping with the ordered set:

7. Mean delay at access,
2. BER end-to-end,
5. Block error rate end-to-end,
15. Time between outages (MTBO),

from Table 19. These perfectly or nearly perfectly matched parameters give no clue on how to discriminate between the two parameter sets. Like the empties of Table 21, the perfect matches can be deleted from further scrutiny. The so deleted rows and columns are accentuated in Table 22.

What remains are the P's. They represent partial relationships, as well as one-to-many or many-to-one impacts, between the two groups. One can distinguish several differences:

Table 22. The Perfect Match Parameters

Primary and Secondary Parameters from FS-1033		Common User Parameters from Table 19															
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Access	1																
	2																
	3	P					P										
Information Transfer - Bits	4								P								
	5																
	6									P							
	7			P													
	8			P													
	14												P				
	16												P				
Information Transfer - Blocks	9								P								
	10																
	11									P							
	12				P												
	13				P												
	15													P			
	17													P			
Diseng.	18																
	19																
Outages	20																
	21	P															P
	22	P										P					

(1) Finer resolution by FS-1033

The common parameter 3, Bit count integrity stands for at least two FS-1033 parameters. They are 7, Bit Loss Probability, and 8, Extra Bit Probability. Likewise, common 4, which deals with integrity of blocks, refers in FS-1033 format to 12, Block loss probability, and 13, Extra Block Probability. The common 8, Mean delay end-to-end, can be broken into two FS-1033 parameters. They are 4, Bit Transfer Time, and 9, Block Transfer time. The common parameter 9, i.e., Misdelivery probability, again corresponds to two FS-1033 parameters, namely, 6 (for bits) and 11 (for blocks). The common 12, Throughput rate for bits, is divided in the standard into 14, Bit Transfer Rate, plus 16, Bit Rate Efficiency. Finally, the common parameter 13 from Table 19, called Throughput rate for blocks, is similarly divided by the FS-1033 into 15, which deals with Block Transfer Rate, and 17, which deals with Block Rate Efficiency.

If the finer solution of FS-1033 is not needed in certain military applications, then the 16 common parameters would appear to be sufficient - at least from this particular partial (P) point of view.

(2) Vagueness of common parameter interpretation

Parameter 1 in Table 19 is entitled Availability of service. This term can have various interpretations and definitions. According to FS-1033 terminology it can be interpreted, at least partly, as 3, Access Denial Probability, or a composite of 21, Outage Duration, and 22, Outage Probability.

(3) Overlap of common parameter interpretations

There are three cases where the meanings of common parameters tend to overlap. In Table 22, for instance, the common parameters 1, Availability of service, and 6, Blocking/delay probability (GOS/QOS), both deal with chances of getting prompt enough access. It is not clear how one draws a distinction between the two. In FS-1033 format, they both would be represented by parameter number 3, Access Denial Probability. Likewise, common parameters 1, Availability of service, and 16, Time to service restoral, are somehow related. Under the FS-1033 standard, their turf would be covered by 21, Outage Duration. And finally, the common parameters 1, Availability of service, and 11,

Outage probability, both reflect the FS-1033 parameter 22, Outage Probability; if not fully, then in part.

The analysis of voids, X's, and P's in Tables 20 to 22 suggests that in several respects the FS-1033 is equal to or preferrable over the common, user oriented, digital performance parameter set given earlier in Table 19. But the proof of any tool is in its actual use. In the next sections of this report, the Interim Federal Standard 1033 will be put to a practical, AUTODIN-II type test. Guided by the parameter definitions of this section, by Appendix B, and by references, quantitative parameter values will be assigned and discussed for the selected digital service mode.

5. ASSIGNMENT OF NUMERICAL VALUES TO THE DIGITAL SERVICE MODE

5.1 Introduction

This section attempts to complete the statistical parameter selection by focusing on FS-1033, and by assigning numerical values to the digital service parameters selected. This being an initial attempt, only limited types of AUTODIN II transactions are treated (McClary, 1977). Main attention is focussed on the interactive (I/A) class of the 56 Kb/s Mode VI service (see Section 3.3). But, before one becomes engrossed with the specifics of parameters and their numbers, it seems advisable to take a broader look at the nature of the number assignment itself.

Here, as noted before, one is concerned with user-oriented performance measures. The numbers, whatever their values, depict the performance expected by the users at the user-system interface. If the actual delivered service results in poorer performance numbers, several things may happen. The offered service may be entirely unacceptable. In this case, the user may be forced to seek other service options, perhaps with other systems.

The offered service may be acceptable, but impaired below the expected standards. In this case, the service clearly loses some of its utility, perhaps differently for different classes of users. The perceived service value, or the degree of acceptability or unacceptability, depends on the user-oriented credibility of the performance numbers. It is thus quite important that the numbers be meticulously determined to represent the real world needs of major military user categories. Whenever operational statistics are available, they should be used. Experimental tests on interactive man/machine communications may constitute valuable guidelines. In other cases, documented military requirements may have to be either substituted for or collated with empirical statistics. Finally, of course, numbers naturally gain in acceptance and confidence if they agree with reality and common sense. This is particularly true when projected numerical AUTODIN II service values are compared in a general way with numbers observed by users of other existing networks. Networks, such as ARPANET, Telenet, TYMNET (Schwartz, et al., 1972; Roberts, 1978), plus the digital services offered by the various common carriers, appear at first glance to be a good source for numerical data. Unfortunately, there are two problems. First, user requirement, satisfaction, or preference numbers have seldom if ever been gathered for the above network services. And second, one suspects that being driven by similar technologies, the performances of the different systems may be too similiar. Major user classes might very well have become accustomed and adapted themselves to the performances offered. It may thus be invalid to infer from the uniformity of system performances that user needs are also the same everywhere.

Growing user needs may be met in part by new and developing technologies. In the case of AUTODIN II, where advanced communications and computer art is to be integrated, the received service will be the result of the latest and most expedient hardware, software, and systems architecture. Performance parameters should not be specified so extreme as to be unrealizable or marginal with the DCS functional designs of the 1980's. At the same time, the technological capabilities should not be wasted or ignored. Instead, system designs and resources must be deployed efficiently and cost effectively to realize the postulated AUTODIN II end-user performance numbers.

The general relationship between network design and service is illustrated in Figure 26. The AUTODIN II service classes, categories or modes, and their respective performance needs are of prime concern here. The received performance clearly depends on the implemented system capabilities. Operational capabilities are limited by several basic constraints. Besides the obvious cost factor, they are affected by traffic volumes and related statistics, such as local and network-wide busy hours, peak factors, nonrandom arrivals, excessive message or transaction durations, and so forth (Rudin, 1974). Other constraints with which one must reckon are the geographical dispersals of subscriber locations, as well as fixed sitings for major network nodes. The distances between sites must be spanned by either terrestrial or satellite circuits. Channel costs and tariffs typically increase with mileage. Expenditures can be alleviated through multiplexing and concentration, but the latter can cause blocking, delays, and even message losses, under certain circumstances. To the extent that system constraints result in different operational and performance characteristics, the constraints comprise a significant factor in the design-for-service picture of AUTODIN II. However, most of these are issues to be treated in future programs. Here one emphasizes the service and performance.

Most constraints to service and performance can be overcome by expansion or enrichment of system design. The parametric performance improvement, so obtained, is typically bought at increased systems cost. In practice, cost is a key ingredient that too often transcends the issues of technical system design, constraints and services. Funds are nominally expended to receive some net benefit (Abrams, 1974; Gitman and Frank, 1978; Licklider and Vezza, 1978). Thus, be it reduced access time, improved BER, or any combination of parametric performance improvements

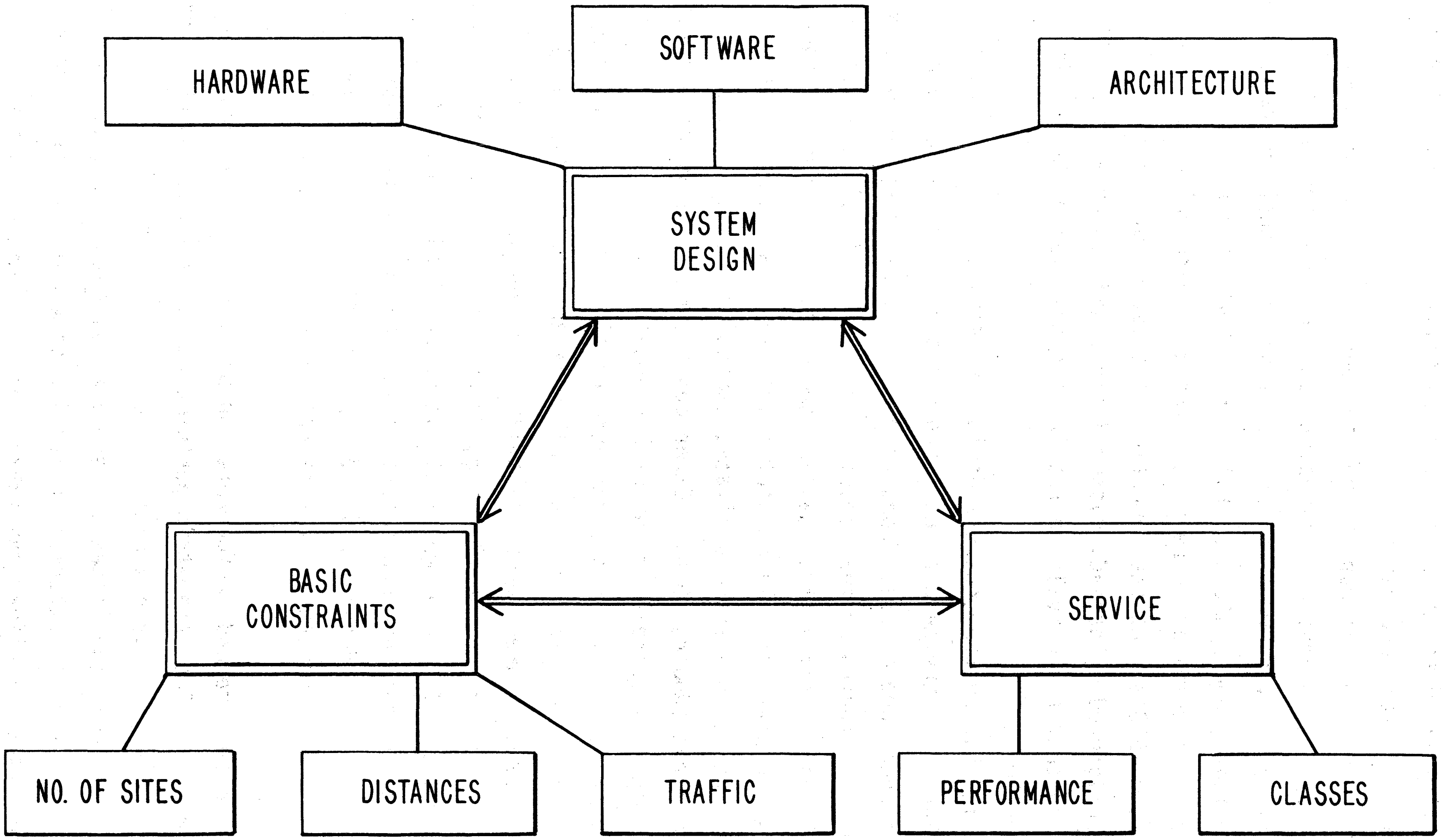


Figure 26. Network design for service needs faces several basic constraints.

to the user, some system cost is associated with said improvements. The situation is briefly sketched in Figure 27. Note that performance improvement represents some benefit or value to the service user. Different service user classes may realize different net value gains. The work of Streeter (1972, 1974) deals with such benefits and costs, but only for scientific computer applications. Similar extensions to the main military computer communications activities appear needed. Figure 27 is, of course, to be interpreted in a very general and qualitative way. All comparisons are relative. The AUTODIN II system engineer, nevertheless, can conceivably use this figure as follows. Parametric performance values are to be selected near the center of the working region, i.e., around the middle or optimal part of the interval where the value received exceeds the system cost outlays. In the case of several service or user classes, an appropriate weighted sum of the individual value curves produces a single value-realized curve. The latter situation applies to AUTODIN II. It is unfortunate that such quantitative value vs. cost curves, at least to these writers' knowledge, are not available for AUTODIN II. Their presence would render the numerical selection process relatively straightforward and deterministic. As it is, one must resort to considerably vaguer, more qualitative, common-sense approaches that often rely on the available system oriented numbers (Cole, 1972; Grubb and Cotton, 1978).

5.2 The Selected Mode Characteristics

Section 3.3 has described in detail the selection of the preferred digital service mode for this study. As noted, the 56 Kb/s Mode VI has been selected for user oriented parameter quantification. The chosen mode is a synchronous, binary, full duplex (FDX) mode with Advanced Data Communication Control Protocol (ADCCP) (Green, 1979). This high data rate mode serves host computer installations. Lower data rate versions of Mode VI are used by various intelligent terminals as well (see Figure 25). More than a dozen installations of the Mode VI (Sevcik, 1977), 56 Kb/s data rate, are projected for a major AUTODIN II Packet Switching Node (PSN), such as Andrews.

The transactions between subscribers (e.g., host computers) and the network PSN's consist of formatted segments of data and a prescribed protocol of interaction. All user systems are asked to conform to the segment format and protocols.

Physically, the segment is a packet, a basic entity, or a block of bits, used for communications exchanges. Segments are used by PSN's to communicate with Mode VI binary subscribers. The interactive (I/A) or Class A version of the Mode VI segment is shown in Figure 28. Central to the segment is the binary text field.

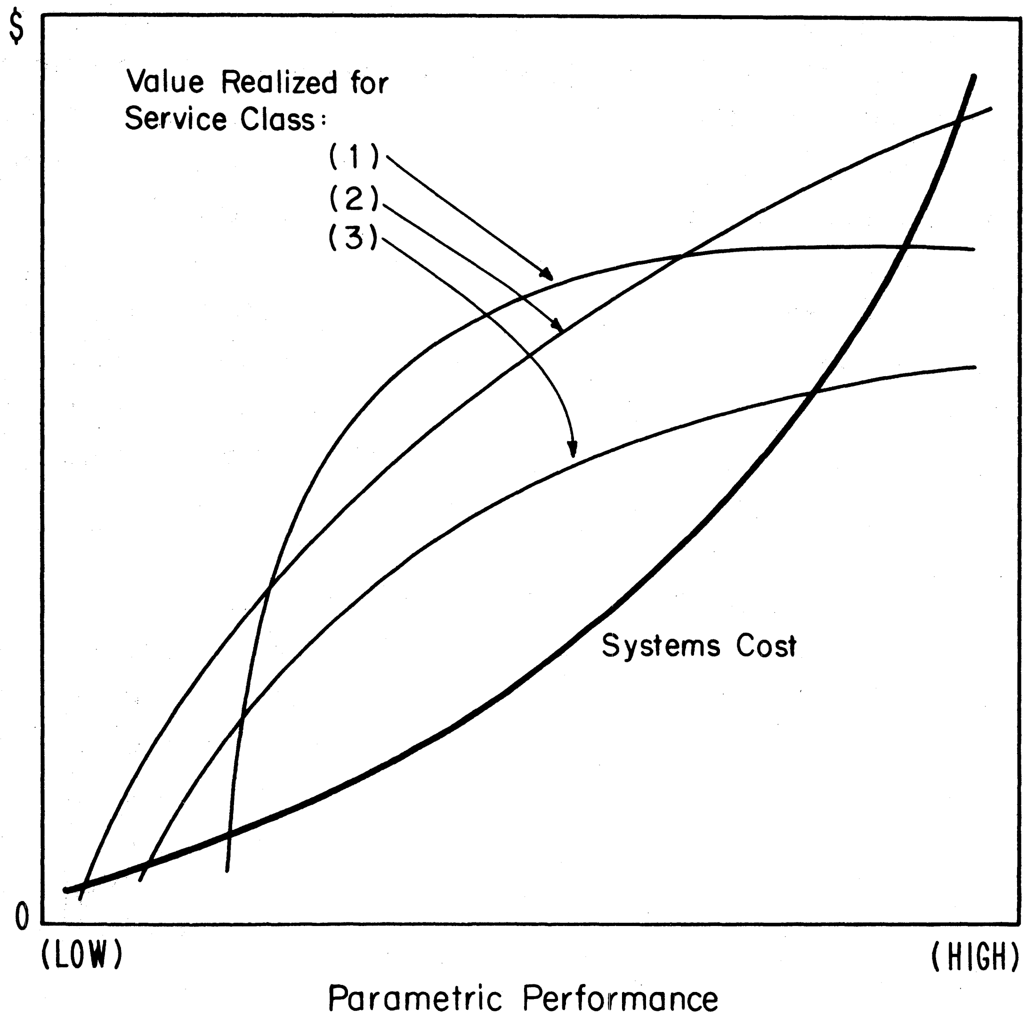


Figure 27. The effect of generalized parametric performance on the value vs. cost decision region.

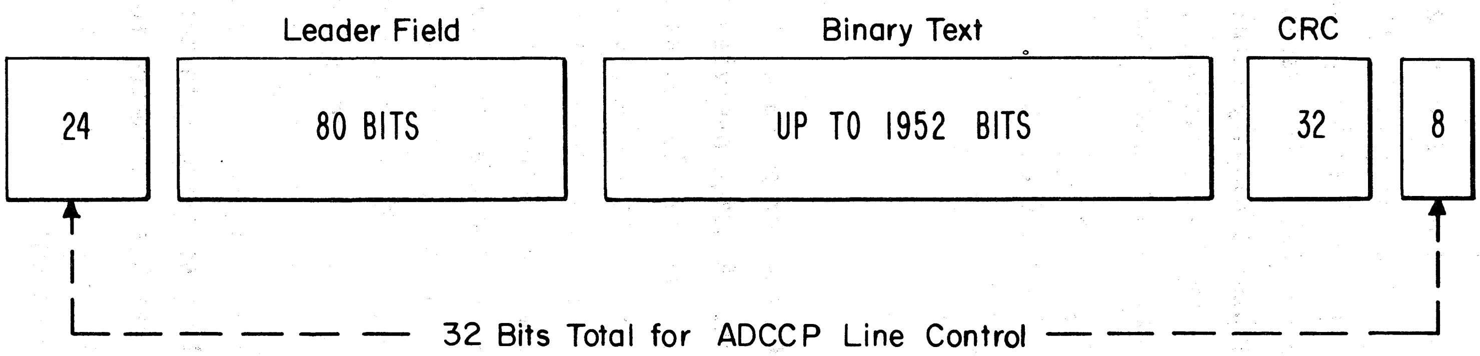


Figure 28. A 2096-bit version of the AUTODIN II Mode VI segment.

The length of the binary text is variable, not exceeding 1952 bits. It is preceded by the so-called Binary Segment Leader (BSL). The BSL length is a constant 80 bits. The binary text is followed by 32 Cyclic Redundancy Check (CRC) bits, also called parity check bits, or the Frame Check Sequence (FCS). This is a departure from the 16 bit CRC field used by ADCCP in the past. The CRC field serves error control purposes through error detection and Automatic Repeat Requests (ARQ).

The very initial 24 bits of the segment, as well as the final 8 bits, are used for line control. The total modified ADCCP field in Figure 28 amounts to 24 + 8 = 32 bits. Other, 40 bit total, ADCCP variants are possible. For practical purposes, the total segment length is roughly equal to or shorter than 2100 bits.

The Mode VI traffic serves many applications and can be variously categorized. The following transaction categories are used often:

- Interactive (I/A) - Class A.
- Query/Response (Q/R) - Class B.
- Narrative - Class C1.
- Bulk 1 - Class C1.
- Bulk 2 - Class C2.
- AUTODIN I.

In terms of typical numbers of bits per message, I/A transactions tend to be the shortest and Bulk 2 the longest. A quick overview is provided in Figure 29. Here the number of bits per message are plotted as abscissa and the typical system response or message delivery time (i.e., access plus transfer time delays) as ordinate. The response time is more appropriate for the I/A services. It is defined (Kelley, 1977) as the time interval between the last user event (e.g., character) and the first system response event (e.g., character). There is considerable overlap of transfer time and message length regions, and their boundaries are by no means sharply defined. AUTODIN I category is purposely not shown in Figure 29. Being roughly in the middle, both in message length (typically 10^2 to 10^6 bits) and in transfer time (typically 10 to 10^4 seconds), the AUTODIN I traffic overlaps the center of the figure. This does not alter the main point that the I/A traffic is usually associated with the least user-to-user delivery time. Of all the groups, I/A has been estimated to constitute from 23% (DCA, 1975) to 50% (Kelley, 1977) of the busy hour traffic. In this sense I/A is the largest AUTODIN II transaction category.

It will be noted below that the interactive category consists of a sufficient mix of application subcategories, traffic acceptance, and criticality classes, to

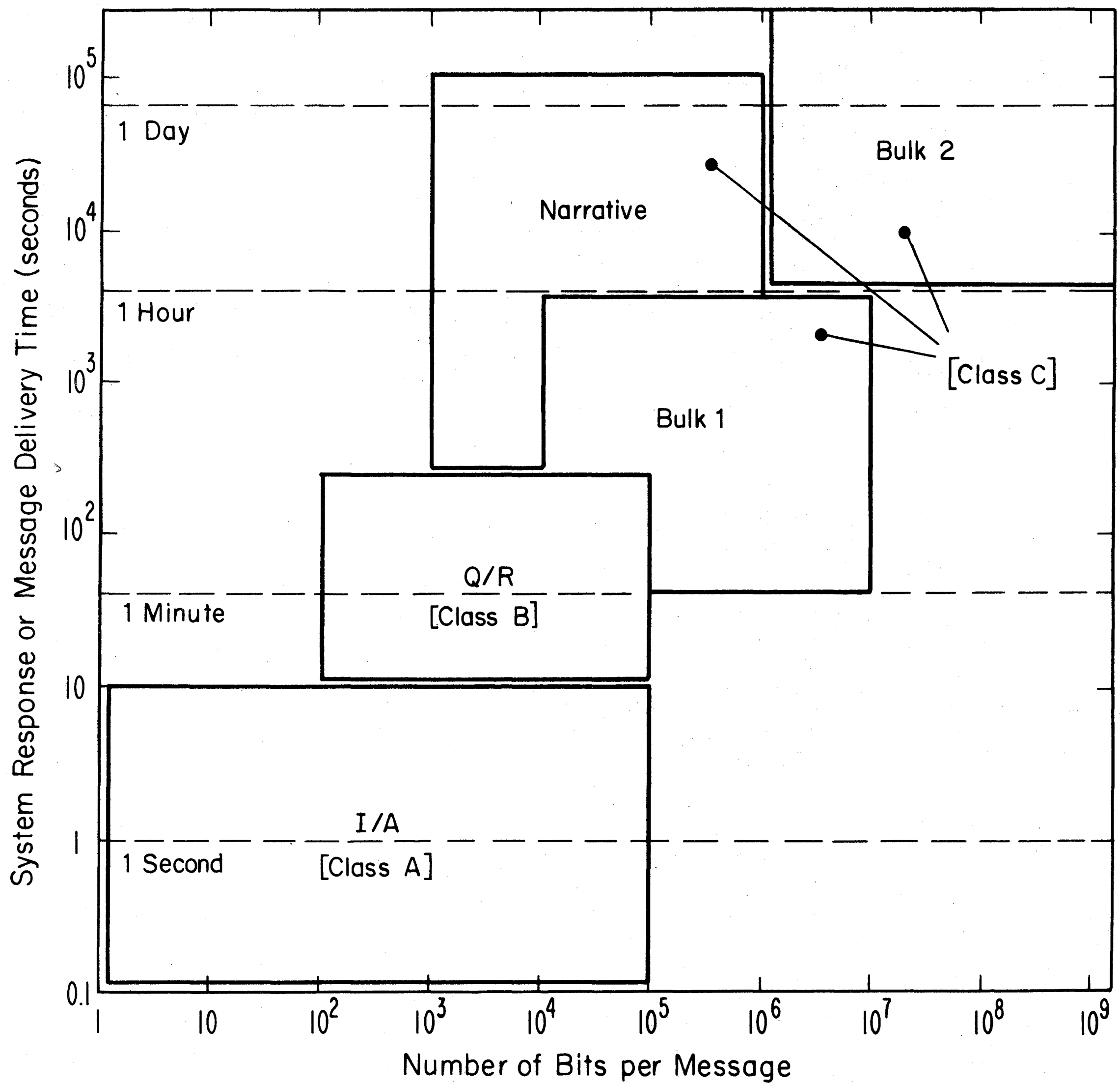


Figure 29. Message lengths and approximate transfer time ranges for the main AUTODIN II application categories.

require generally similiar efficiency, accuracy, and reliability numbers, with one exception. That exception is the message delivery time requirement. It appears to be the most stringent for the I/A, Class A, transactions. For that reason, the interactive (I/A, or Class A) traffic has been selected as the category to be studied first under this numerical assignment task.

The delays associated with the I/A traffic fall under Efficiency in the FS-1033 26-parameter field of Section 4. The pertinent delay times are called Access Time, Bit Transfer Time, Block Transfer Time, and Disengagement Time. Reduced transfer times for I/A are possible because of the continuous session setup for I/A traffic. While both I/A and Q/R transactions may often be of comparable length, the session continuity or its absence, makes a difference. The Q/R exchanges make no effort to sustain continuity. Every transaction exchange establishes its own independent path through the system. In I/A, on the other hand, a series of transactions between two interacting subscribers are grouped into continuous sessions. The access and message transfer times are nominally largest at the beginning of a session. Within the session, however, due to the dedicated path (plus facility) nature of the session, the delays are reduced (Sanders, 1980).

Several application subcategories constitute the I/A category for the 56 Kb/s Mode VI. They include all precedence levels:

- Flash Override (Y & W).
- Flash (Z).
- Immediate (O).
- Priority (P).
- Routine (R).

For I/A traffic acceptance purposes, these levels are further specified as being either nonblocking or blocking. Nonblocking traffic has preemption capability over all other traffic. The nonblocking traffic is said to belong to Traffic Acceptance Category I (more specifically, I4 for Class A interactive mode). It is composed of Flash Override (Y & W) and Flash (Z) precedences. Blocking categories are numbered II, III, and IV, and they consist of Immediate (O), Priority (P), and Routine (R) traffic, respectively. Of the entire I/A traffic, only about 1% belongs to Category I (DCA, 1975). Roughly 15% belong to Category II, 38% to Category III, and some 46% to Category IV. The 56 Kb/s part of I/A is expected to be similarly divided. The interactive Mode VI, 56 Kb/s, service applies almost exclusively to either single or multiple logical channel host computer transactions. The only possible exception appears to be the graphic/light pen application of the RAND Tablet Terminal. The light pen appears to meet all the Mode VI prerequisites, namely ADCCP, CRC, 56 Kb/s, and I/A. It also has a cross connection requirement to Mode VI computers, but it is not quite certain whether the pen is actually classified as part of Mode VI.

When a host computer communicates through the AUTODIN II network, its correspondent terminal at the other end may be one of several entities. It may be one or more computers. Or it may be a device used for controls, alarms, status indication, monitoring, telemetry, and the like. The terminal devices can also accommodate some human intervention or interaction. The latter appears, however, as a limited application for the 56 Kb/s I/A mode in the near-term future. The typical subcategories of the Class A interactive traffic are illustrated in Figure 30. There is one subcategory involving potential human interaction. It evokes the longest response time, perhaps up to 15 seconds. The two other subcategories involve only machine processes. The "short" message subcategory consists of messages whose length does not exceed 10 bits. The "long" message category can extend beyond 10^5 bits per message.

5.3 Access Phase

Three FS-1033 parameters specify the access performance as seen by the user. As noted in Table 18, the access efficiency or speed is measured by parameter (1) Access Time. Access accuracy is measured by parameter (2) Incorrect Access Probability, while the reliability of access is depicted by (3) Access Denial Probability. In proposing AUTODIN II candidate numerical values for these three parameters, several previous assumptions are used. Most significantly, the parameters are nearly the same as defined in FS-1033 (GSA, 1979; Seitz and Bodson, 1980, or Appendix B). Minor modifications, such as additions, are made only to suit apparent military communications needs. All parameters are user-oriented service performance descriptors, which are largely independent of the system. The numbers reflect primarily user service requirements at the user-system interface and not system implementation, or capabilities or limitations of any specific system. This does not, however, preclude the use of these service numbers in system planning, engineering, or implementation later on. And finally, the postulated service is of the previously introduced 56 Kb/s, Mode VI, interactive (I/A) type, as described in Section 5.2.

5.3.1 Access Time

Access Time is defined in FS-1033 as the average value of the elapsed time between the start of an access attempt and the realization of a successful access.

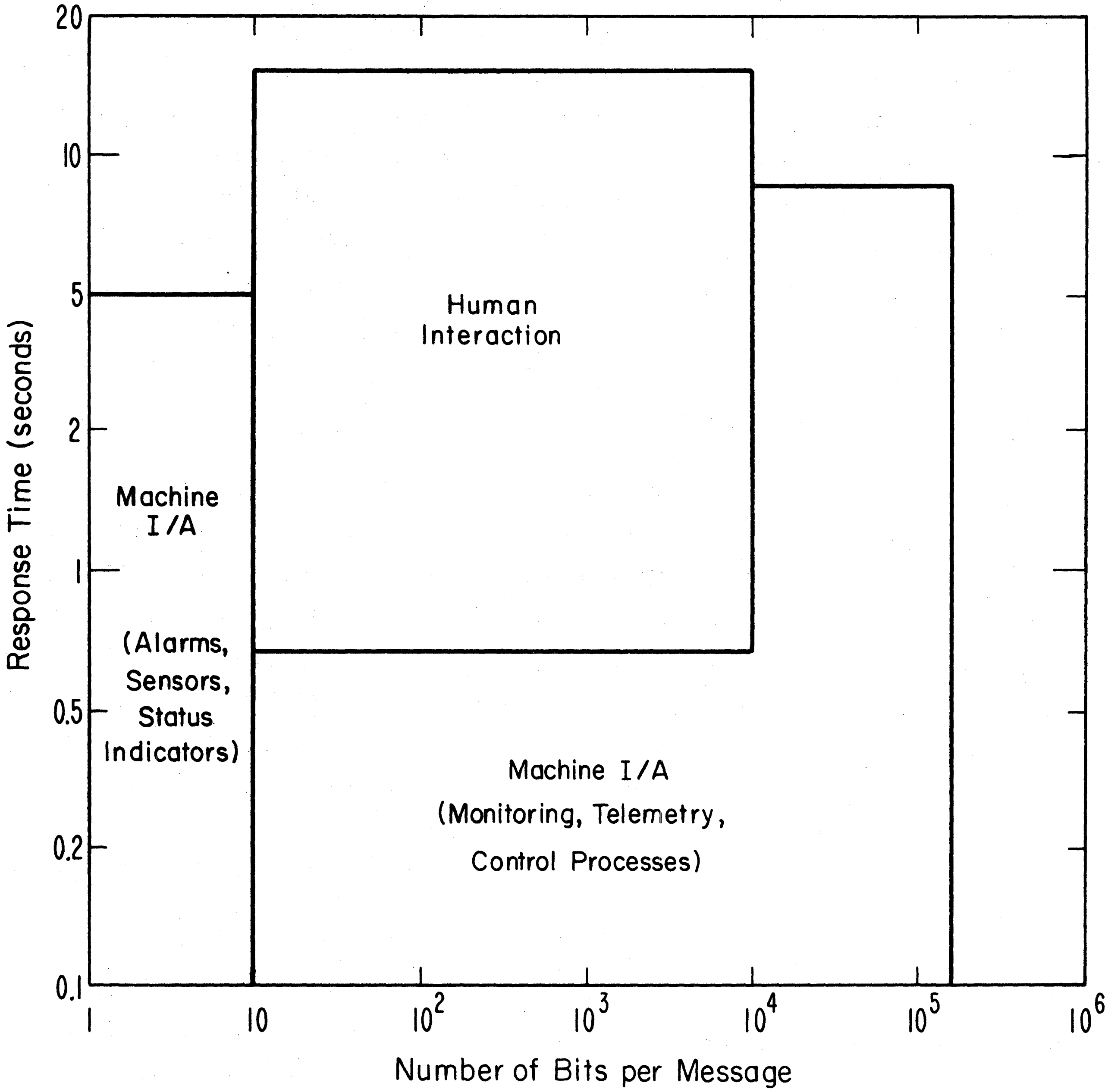


Figure 30. Typical subcategories of the interactive, Class A, traffic.

As such, the definition applies individually to each user-to-user pair. Access interval ends when the first bit of source user information is entered through the user-system interface into the system. Only successful accesses count toward this elapsed time measurement. Access is defined as unsuccessful if it is not concluded within a specified maximum time interval, also called time out.

In virtual-circuit systems, such as for interactive transactions on AUTODIN II, it is necessary for the intended destination host, or its front end emulator, to be contacted and committed to the I/A session. This initial session commitment may be a major contributor to larger access times. There are other delays at the local, distant, and tandem switching nodes (Sanders, 1980). The main features of the Access Time profile are illustrated in Figure 31.

Finally, one should not overlook the fact that the Access Time variability on individual user-to-user paths is but a part of a larger statistical problem. FS-1033 recommends time series averaging for the individual user pairs. It is possible, however, that distinct user-to-user pairs may generate widely different access time distributions. A lot depends on network layout, node and link capacities, and the traffic flow through the network. This variation must be superimposed on the random time series observed at a fixed user-to-user pair. Because of estimated 50-100 Mode VI, 56 Kb/s AUTODIN II terminals (DCA, 1975), the number of distinct user pairs could be in the thousands. Some kind of grouping or averaging of access times must be administered to render the performance description and planning more tractable. Figure 32 indicates a two-fold grouping assumed here. The user pairs are grouped according to I/A application (see Figures 29 and 30) and according to service acceptance criticality (or precedence). This structure offers as many as 3 x 4 = 12 different service categories.

The Access Time values to be specified here are intended for the busy hour and are uniformly administered or observed over the CONUS service area.

A minor modification of the FS-1033 definition appears warranted. As noted, the Interim Standard emphasizes the average or mean value of the Access Time. Unfortunately, in the majority of both military and nonmilitary applications there is also considerable interest in seldom occurring performance values far worse than average performance values (Miller, 1968; Rose and O'Keefe, 1980). Both users and system engineers are often concerned with averages and dispersions about the averages. Some applications must avoid marginal or unacceptable operation. That is the extreme bad end of the performance, such as Access Time, distribution. The region typically includes distribution function, i.e., probability of being less than or equal to, values of 0.9, 0.99, or 0.999. In terms of percentiles, these

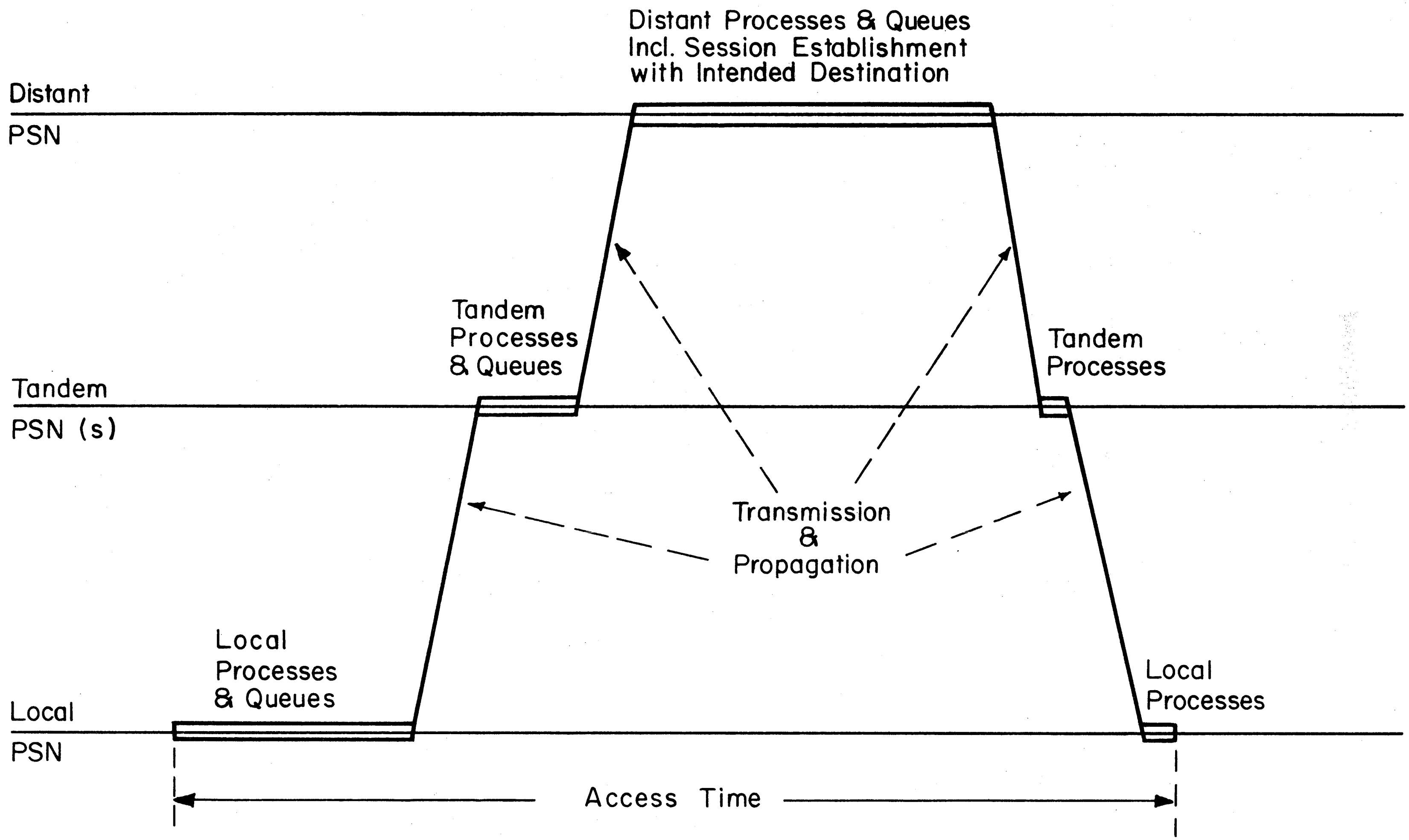


Figure 31. Main delay elements of the Access Time parameter.

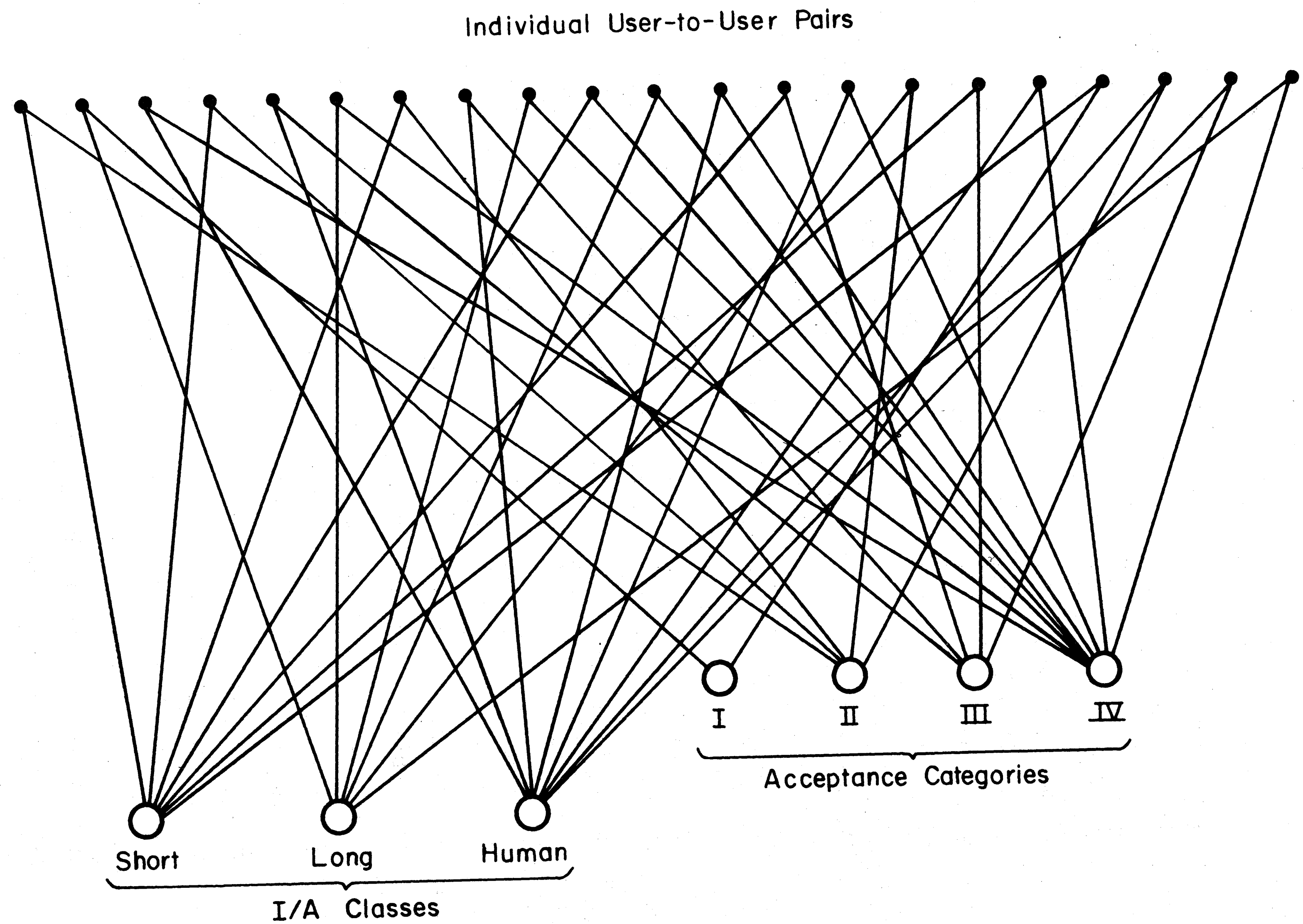


Figure 32. Grouping of user-to-user pairs for performance planning and averaging purposes.

are referred as 90%, 99%, or 99.9% levels. The probability that Access Time, picked at random from the appropriate sample space, will exceed these levels is 0.1, 0.01, or 0.001, respectively. To assess the user oriented performance realistically, the means and medians of the performance parameter can be augmented with percentile values interest.

In what follows, the 90% or 0.9-percentile value is delegated as the supplementary parameter for access time. Unless confusion could occur, the mean value will be simply called the Access Time. Its 0.9-percentile value will be so identified. The 99.9% or higher percentile values are not included in this section. They will be treated in Section 5.3.3, where these play a natural role in the definition of Access Denial Probability.

Because of the scarcity of empirical user service requirement data, the assignment of numerical values must be treated as an example. Furthermore, the pertinence and validity of these numerical values should be established in an appropriate AUTODIN II validation program. The end result here is a listing of performance numbers that fits the meager experimental data base and that is reasonably self-consistent. In what follows, the final result (see Table 23) is presented first. Thereafter, the reasons and explanations of these numbers, plus various user-system implications, are examined at length.

Candidate Access Times for the interactive (I/A) 56 Kb/s, Mode VI, service are proposed in Table 23. As noted in Figure 32, there are three application categories and four precedence categories. For each of the 12 classes, the table proposes the mean value plus a 0.9-percentile value for the Access Time. Within each class, the requirements and facilities are assumed to be uniform and with less variance than between classes. The numbers given are not based on new empirical work, such as statistical user surveys. The numbers reflect consensus interpretation of pertinent published work and DOD documentation.

The columns of Table 23 represent the four traffic acceptance categories, from Flash Override to Routine. The three main rows divide the I/A traffic into the same three subcategories, as noted earlier in Figure 30. The fully machine-implemented I/A subcategory with no more than 10 bits per message is denoted as "short." The machine I/A subcategory with 10 to 10^5 bits per message is called "long." Finally, the third I/A application category experiences occasional human intervention or override.

The human reaction times have their minimal physical limits (Boies, 1974; Elam, 1978). Nominally, the limits cluster around 0.5 seconds. The only Access Time value shorter than that is proposed for precedence I, human I/A, traffic.

Table 23. Proposed Access Times for Twelve I/A 56 Kb/s Mode VI User Categories

Applications	Level	Traffic Precedence Category			
		I(Y,W,Z)	II(O)	III(P)	IV(R)
Machine I/A Short	Mean	0.1*	0.2	0.4	0.75
	0.9-Percentile	0.15	0.3	0.6	1.0
Machine I/A Long	Mean	0.2	0.4	0.6	1.0
	0.9-Percentile	0.25	0.6	1.0	2.0
Human I/A	Mean	0.3	0.6	0.9	2.0
	0.9-Percentile	0.5	1.0	1.5	4.0

*All numbers in seconds.

All other human I/A categories have higher mean and percentile Access Time goals, the largest being 4.0 seconds for the Routine (category IV or R) traffic. The effect of user dependence should be further examined, especially for the human interaction numbers. For instance, manual entry of destination addresses would take considerably longer.

Automatic equipment can react much faster than humans. However, even in the high speed host computer and programmed terminal application, there are limits on how fast Access Times need to be. Judging from the literature on user needs and preferences (Martin, 1972 and 1973; Kelley, 1977), response times lower than 0.1 second are not required by the interactive machine users. In Table 23, only the short message class is shown to have a mean that low. For precedence I, the short messages call for 0.9-percentile time of 0.15 seconds. Such short times appear difficult for several reasons. First, they are impossible for satellite circuits. And second, the conventional definition of response times differs from the FS-1033 Access Time. The messages from the long machine message class take a longer time to conclude their transactions. It appears reasonable to relax the Access Time goals for the long machine messages relative to the short ones. On the other hand, even the long machine messages should be accessed quicker than those with a human hand at the controls. In Table 23, the long machine message Access Time requirements are given roughly as half-way between the two extremes.

Available user-oriented service numbers can be used in support of proposed values. Such rationalization is done in Figures 33 to 35. The results quoted are those of Martin (1972) and Kelley (1977). It should be emphasized that these results represent "response" times. The response time definition resembles that of system delivery times within an established I/A session, but is by no means clear in all cases. As noted in previous sections, the vagueness of interfaces and key events makes the response time meaning questionable vis-à-vis the FS-1033 Access Time. Nevertheless, when a system takes a long time to "respond," less and less can be gained by further reduction of the Access Time. Likewise, quick "response" times have potential user benefits when paralleled by comparably small Access Times (Stewart, 1979; Rose and O'Keefe, 1980). Thus, the two time measures should be correlated.

There may be occasions where the 12 value requirement set of Table 23 (same as in Figures 33 to 35) is too elaborate. A single number may be preferred. One obvious choice is the worst-case value, such as 0.1 seconds for the mean and 0.15 seconds for the 0.9-percentile. However, such a choice may be too stringent. One may prefer some representative, weighted average over the 12 Access Times shown. Off hand, it is not clear what weighting approaches should be used.

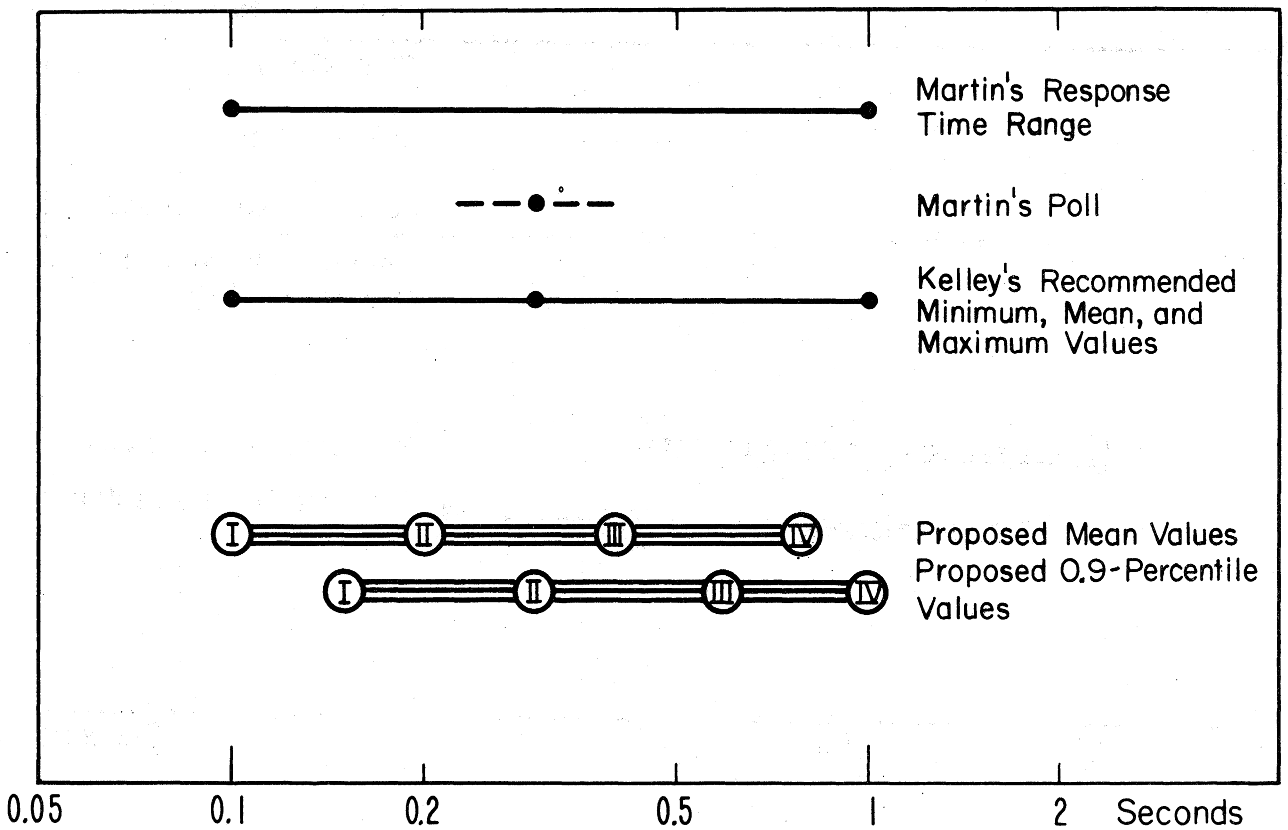


Figure 33. Proposed Access Times for short message I/A machine traffic versus response times.

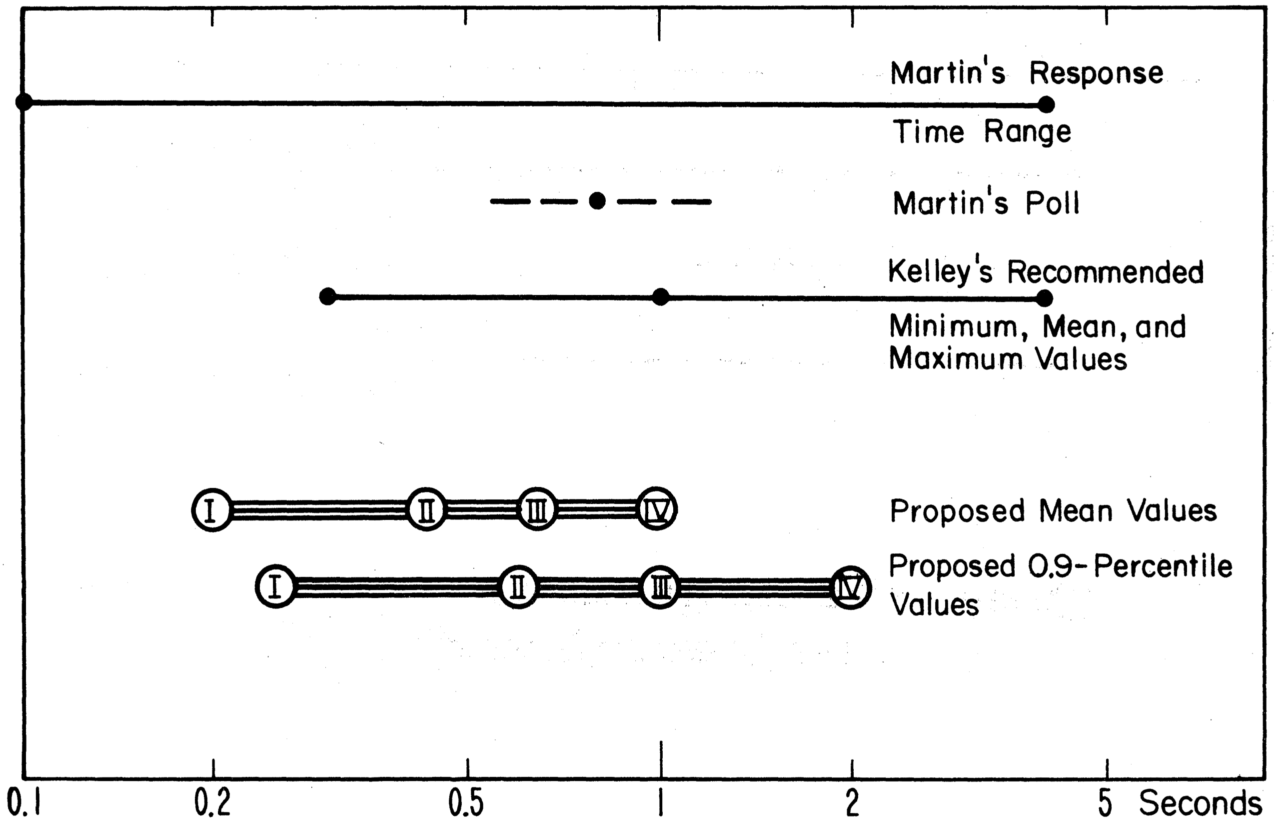


Figure 34. Proposed Access Times for long message I/A machine traffic versus response times.

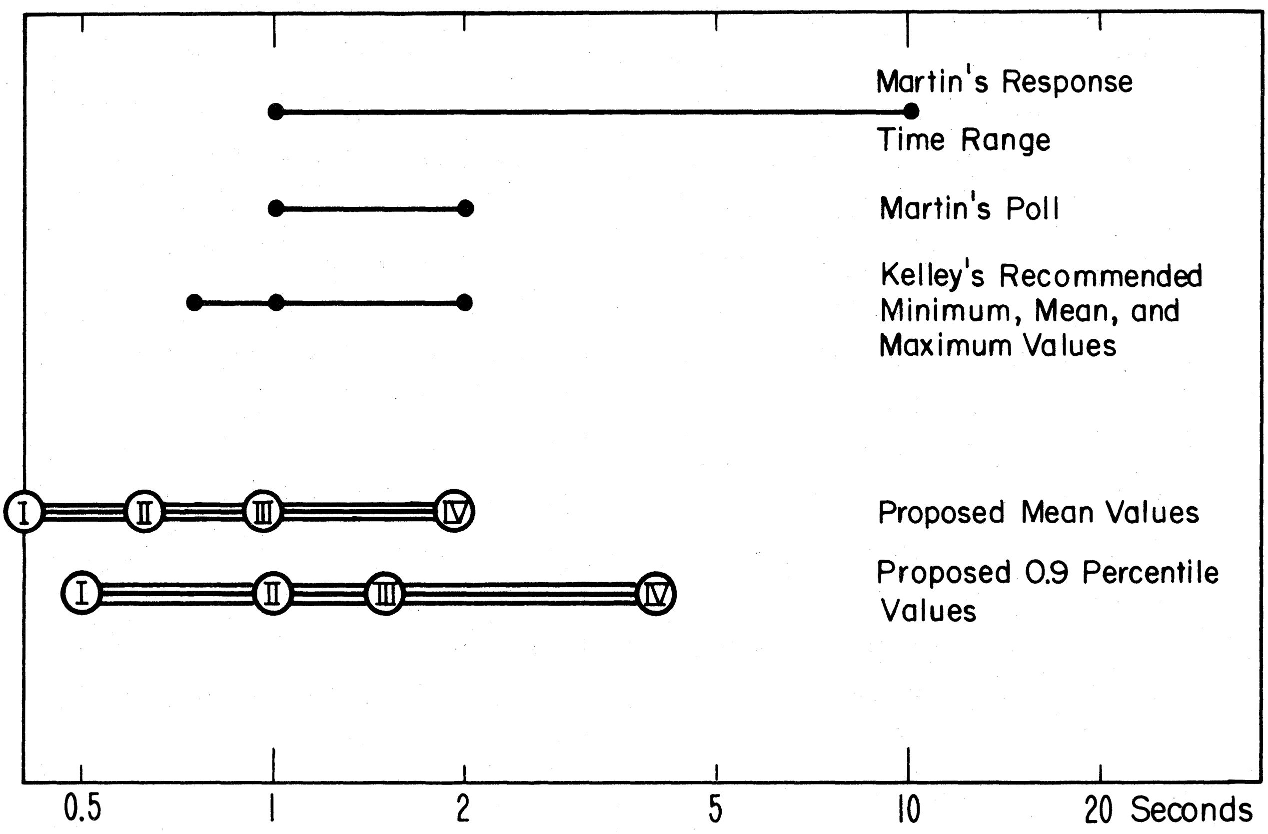


Figure 35. Proposed Access Times for human I/A traffic versus response times.

Messages encounter other network delays after access is completed. If a message contains N data bits, and the effective throughput rate (called Bit Transfer Rate in FS-1033) is R bits per second, then the modem delay alone cannot be lower than N/R. Other, often more significant delays occur at nodes and links. Such functions as packetizing, queueing, error control delays, and general processing take place at nodes. Propagation delays accumulate on both terrestrial and satellite links. Let the total of all node and link delays be DN/R, regardless of the configuration of nodes and links. Then D represents the relative node-plus-link delay in units of modem delay. It is a dimensionless entity. The sum, N(1+D)/R, constitutes the minimal total network delivery delay (equivalent to Bit or Block Transfer Time in FS-1033), that can only be approached when access delays vanish. As is well known (Frank, et al., 1976; Kleinrock, 1976), queueing delays alone can vary appreciably. Since both satellite and terrestrial links can get involved, assume total delay in the 0.1 to 1 second neighborhood. Also assume the effective throughput to be R = 10 Kb/s. Then D values can be estimated for the three different I/A application types.

Entities N(1+D)/R are plotted as straight lines versus N, the number of bits per message, in Figures 36 to 38. For the short machine interactions, the appropriate D values for 0.1 to 1 second delay spreads are around 100 to 1000. These least total delay values are compared in Figure 36 with the proposed Access Time values. In the region of interest, they are of the same order to magnitude, as shown. Neither dominates the response time range for short machine transactions in Figure 36. For long machine messages D can vary around unity. The slanted lines of Figure 37 show the corresponding least total delay for the machine, long-message application category. Finally, the human I/A traffic, with D ranging about 10, is depicted in Figure 38. All three plots support in a rough way the validity of the Access Time numbers suggested in Table 23.

5.3.2 Incorrect Access Probability

Incorrect Access Probability is defined as the ratio:

$$\frac{\text{Number of completed accesses to incorrect destinations}}{\text{Total number of completed accesses}} .$$

Incorrect access can occur in the virtual circuit mode when the entire I/A session is accessed to the wrong distant address. The ratios can be computed separately for individual service subcategories or jointly for meaningful groupings of subcategories.

From the military user point of view, a significant measure appears to be the average time interval that corresponds to the occurrence of a single incorrect

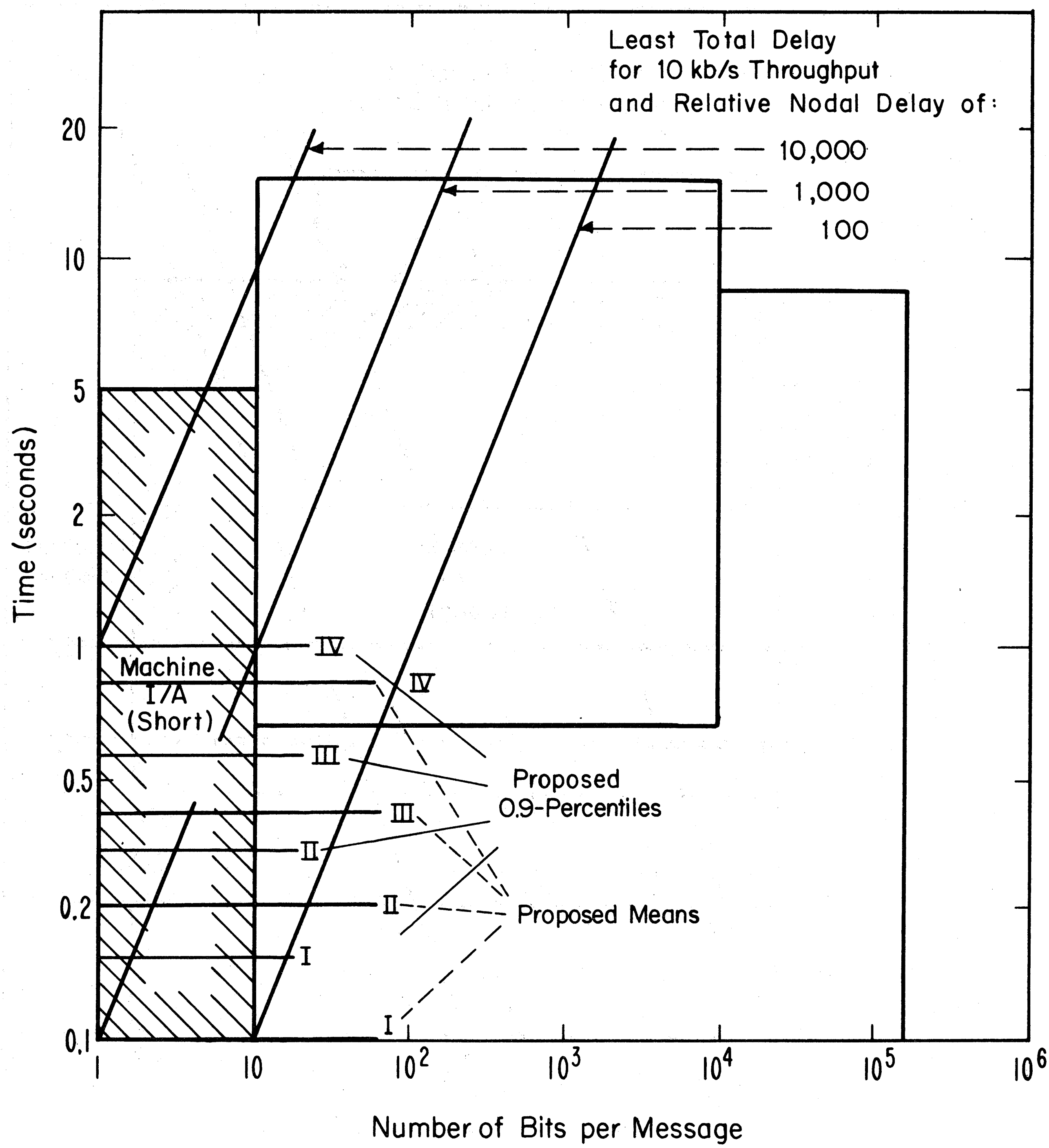


Figure 36. Proposed Access Times for short machine I/A messages versus least total delay.

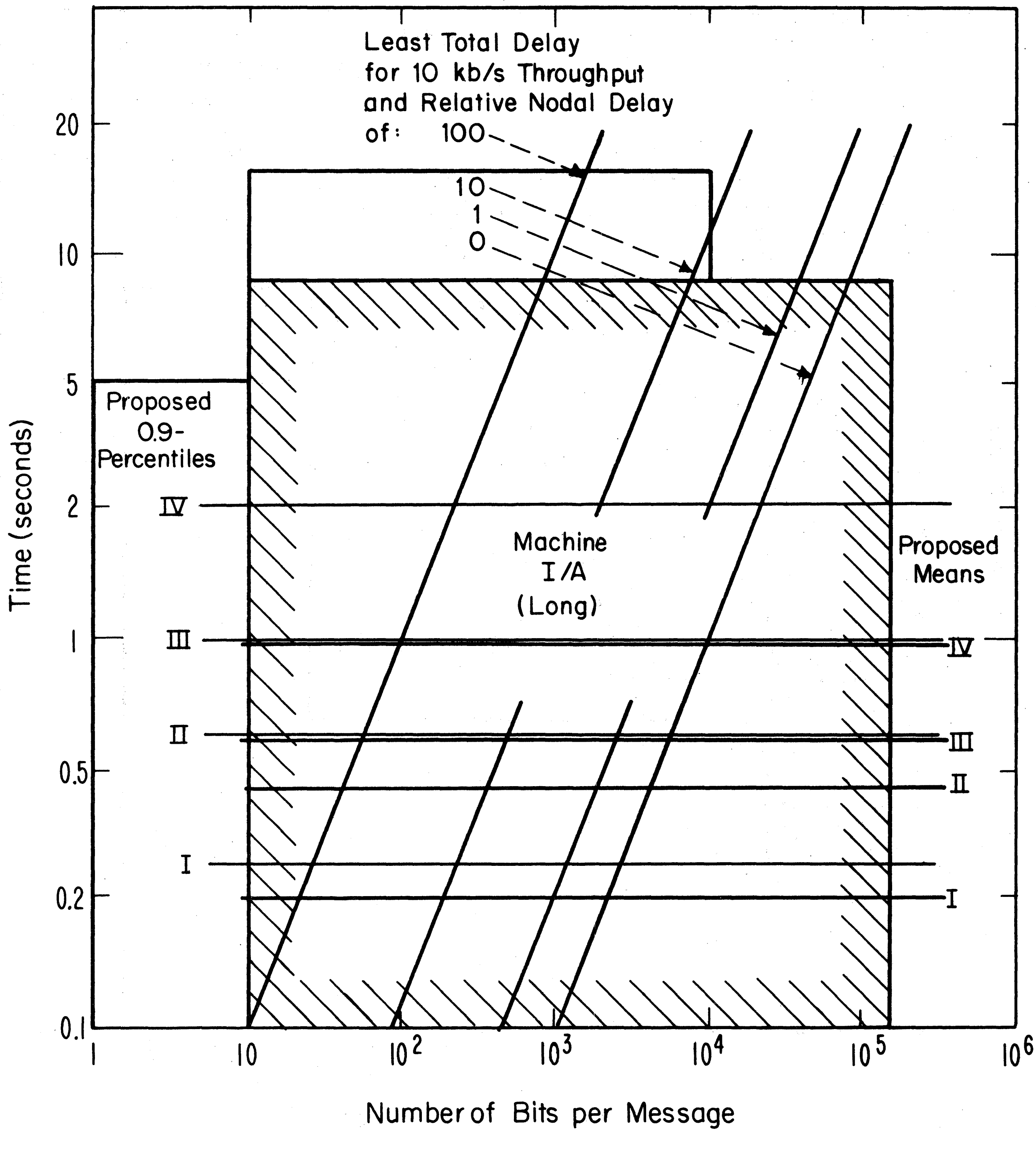


Figure 37. Proposed Access Times for long machine I/A messages versus least total delay.

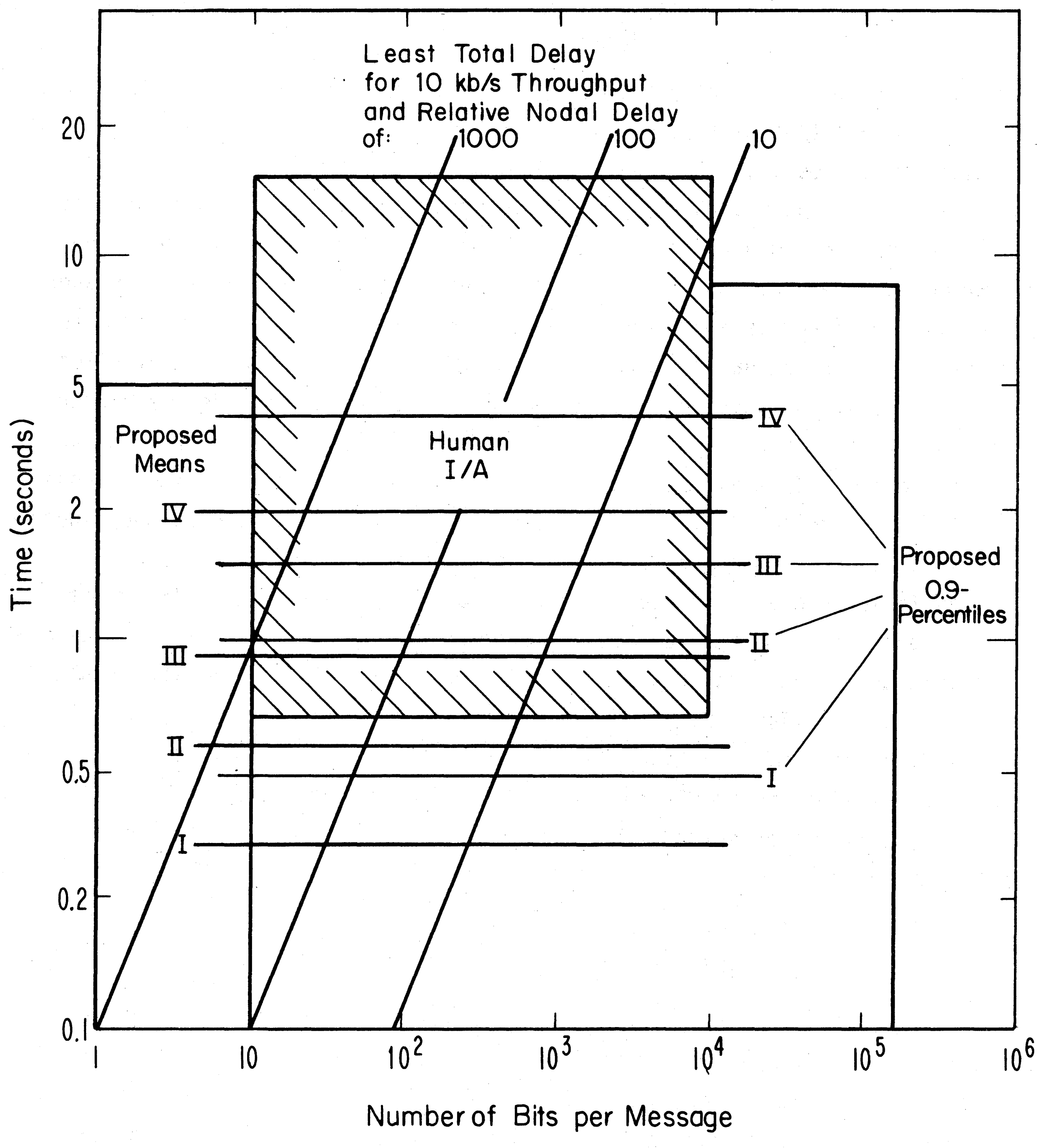


Figure 38. Proposed Access Times for human I/A messages versus least total delay.

access. This average interval can be defined per user, per node, or per suitable traffic categories. In this section, the following systemwide definition is assumed. The average time interval is applied systemwide over the entire CONUS AUTODIN II network, but only to the busy hours (e.g., the potentially worst service hours). This enables one to use documented busy hour (BH) statistics. The average times for a single incorrect access are next presented separately for the three machine and human I/A cases, as well as for two precedence cases: Category I (Y, W, Z) and Categories II, III, IV (O, P, R), respectively.

As a starting point, note that 10^2 busy hours correspond very roughly to a month of operation. Hence, one event per 10^3 busy hours, corresponds to one systemwide incorrect access per year. One event per 10^5 busy hours, corresponds to one systemwide access malfunction in hundred years. A user is expected to require some numerical value in this 10^2 to 10^5 busy hour range for satisfactory operation. Table 24 lists postulated numbers. Note that override precedence level (I) asks for roughly 10 to 100 years of perfect accesses before the next incorrect access. This is about 10 to 100 times longer than is deemed necessary for the nonoverride levels (II, III, IV). The shortest mean time between incorrect accesses is assigned to the human I/A, nonoverride category.

From Table 24 one can infer the values for FS-1033 parameter, Incorrect Access Probability. As before, such values are user service oriented. They apply to the interactive 56 Kb/s, Mode VI, traffic only. One version of the resultant numbers is given in Table 25.

The derivation of these numbers is based on several projections and estimates. Assume, as detailed in DCA, System Performance Specification for AUTODIN II (1975), that the busy hour transmitted traffic for the Mode VI, 56 Kb/s, type over the entire network is:

525,000 kb/hour - for Category I,
1,781,000 kb/hour - for Categories II, III, IV.

In this assumption, Category I traffic has been further enlarged by 100,000 kb/hour to reflect the potential need to utilize the five Flash Override modes that in present documentation reflect 0 volumes for the AUTODIN II scenario (DCA, 1975, Appendix B). Note that this entire Category I volume is on the order of 23% of the total. This particular value is the highest encountered so far. It is considerably higher than the 1% value predicted in Section 3.2.1.2.2 of the same DCA Specification, or in Kelley's (1977) prediction, an issue to which we shall return later. As mentioned previously, estimates of I/A subclass sizes also vary considerably. Interactive traffic may constitute from 23% (DCS, 1975) to 50% (Kelley, 1977)

Table 24. Postulated Systemwide Average Number of Busy Hours per Single Incorrect Access

Traffic Acceptance Category	I	II, III, IV
Precedence Levels	Y, W, Z	O, P, R
Machine I/A (Short)	10^5	10^3
Human I/A	10^4	10^2
Machine I/A (Long)	10^5	10^4

Table 25. Conservative Candidate Values for Incorrect Access Probability

Traffic Acceptance Category	I	II, III, IV
Precedence Levels	Y, W, Z	O, P, R
Machine I/A (Short)	$5.0 (10^{-12})$	$1.5 (10^{-10})$
Human I/A	$6.1 (10^{-10})$	$1.8 (10^{-8})$
Machine I/A (Long)	$3.2 (10^{-9})$	$9.3 (10^{-9})$

of the 56 Kb/s, Mode VI total. The busy hour is alleged to carry the highest relative I/A traffic. In what follows, a 40% I/A component is assumed for all Categories I to IV. Furthermore, the machine (short), human, and machine (long) I/A percentages of the total Mode IV flow are estimated to be 3%, 31% and 6%, respectively (Kelley, 1977). If the average message sizes are, as suggested by Figure 30, around 8 bits, 10^3 bits, and 10^4 bits for the three machine/human classes, respectively, then the number of busy hour messages can be estimated. The resultant numbers are presented in Table 26.

Note that, by an order of magnitude, most transactions belong to the short I/A message classification. The requirement numbers of Table 25 follow directly from Tables 24 and 26. For instance, for precedence I and short machine transactions, one estimates the Incorrect Access Probability as:

$$\frac{10^{-5}}{550 \cdot 60 \cdot 60} \cong 5.0\ (10^{-12})\ .$$

This particular probability number is the lowest in Table 25 and, at the same time, the most demanding one on the system. In the worst case sense, it implies that, as a common user oriented requirement, the Incorrect Access Probability should be on the order of 10^{-12}.

The requirement can be relaxed by several means. In particular, things can be made easier if the Category I (i.e., Flash Override) traffic is reduced from the high 23% assumed in the construction of Table 26. As suggested by Kelley (1977), a suitable number may be around 1%. If one can keep the total I/A traffic constant at 250 Kb/s, then the message rate numbers in Table 26 would be altered substantially. Category I would be reduced by a factor $1/23 \cong 0.043$, while Categories II, III, IV would be increased by a factor $(100-1)/(100 - 23) \cong 1.28$. The main result would be a new set of candidate values for the Incorrect Access Probability value.

The new set is shown in Table 27. Compared to the earlier, more conservative or worst case numbers of Table 25, the new requirement appears far easier to satisfy. Instead of an Incorrect Access Probability level of 10^{-12}, the new easier requirement calls for something around 10^{-10}. By the way, the latter number may be realized in practice by a 32-bit CRC parity check. By the familiar "folk theorem" on worst case undetected block errors, the CRC undetected error rate should always be better than $2^{-32} \cong 2.3\ (10^{-10})$.

Table 26. Approximate Systemwide, Busy Hour, Bit and Message Flow Rates for I/A, 56 Kb/s, Mode VI Substreams

Traffic Acceptance Categories	I (Y, W, Z)		II, III, IV (O, P, R)		Totals	
	Bit Rate (Kb/s)	Message Rate (mess/s)	Bit Rate (Kb/s)	Message Rate (mess/s)	Bit Rate (Kb/s)	Message Rate (mess/s)
Machine I/A Short	4.4 (1.7%)	550 (21.1%)	14.8 (5.8%)	1855 (71.2%)	19.2 (7.5%)	2405 (92.3%)
Machine I/A Long	8.8 (3.4%)	1 (0.03%)	29.7 (11.6%)	3 (0.1%)	38.5 (15.0%)	4 (0.1%)
Human I/A	45.3 (17.7%)	45 (1.7%)	153.3 (59.8%)	153 (5.9%)	198.6 (77.5%)	199 (7.6%)

Table 27. Easier Candidate Values for Incorrect Access Probability

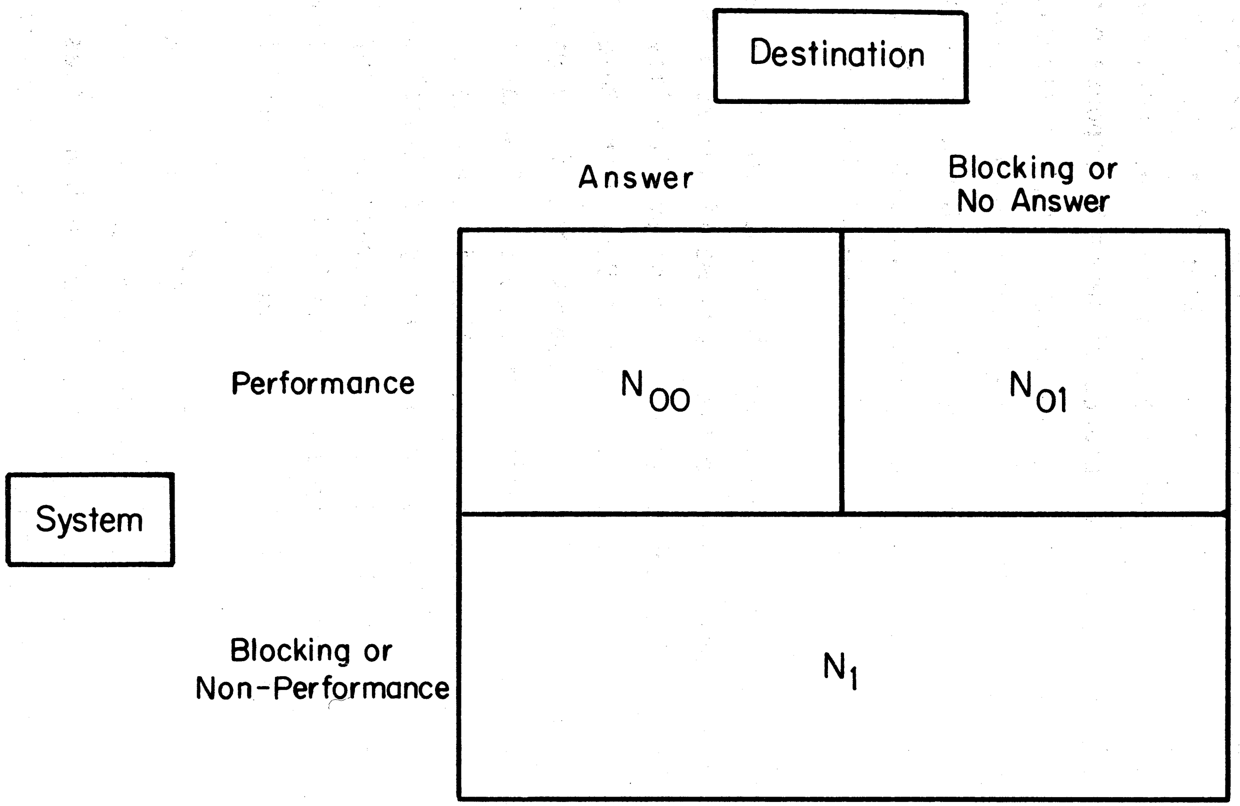
Traffic Acceptance Category	I	II, III, IV
Precedence Levels	Y, W, Z	O, P, R
Machine I/A (Short)	1.2 (10^{-10})	1.2 (10^{-10})
Human I/A	1.4 (10^{-8})	1.4 (10^{-8})
Machine I/A (Long)	7.4 (10^{-8})	7.2 (10^{-8})

5.3.3 Access Denial Probability

Access Denial Probability is defined as the ratio of total access attempts denied by the system to the total effective access attempts. An "effective" access attempt is one that does not encounter destination user nonperformance (failure to answer) or blocking (destination busy). To clarify this definition, the truth table of Figure 39 is presented. Note that when the system performs as needed, the destination user nonperformance or blocking does not enter in the Access Denial Probability calculation. System nonperformance and blocking can manifest itself in two ways: by the occurrence of a blocking (i.e., network busy) signal, and by the absence of an appropriate network response (e.g., no issuance of any meaningful signal by the system). The absence of network response ordinarily results in a nonperformance time-out. The time-out can be defined in several ways. Federal Standard 1033 sets it equal to 3 times the mean (or average) value of the Access Time. When this is done, the time-out values of Table 28 result. One weakness of the numbers given in Table 28 is the lack of numerical data in the published literature that would indicate the probability levels with which said time-outs would occur.

Here, it is proposed to follow a different approach in addition to Table 28. In Section 4, the performance parameter "availability" was mentioned as of significance to military communicators. One can broadly interpret availability as the complement of Access Denial Probability. Sources, such as Buhrke and Mele (1974), Kimmett and Seitz (1978), and Feldman et al. (1979), suggest that for various classes of circuit switched services availabilities in excess of 99% may be unrealistically high. Thus, even in the absence of call blocking, Access Denial Probabilities around 10^{-2} may be typical of existing lower speed digital services. However, the critical nature of the AUTODIN II Mode VI, interactive 56 Kb/s service, plus its advanced technology, infer considerable availability enhancement. Access Denial Probabilities in the 10^{-3} to 10^{-4} range should be explored for the selected mode.

Access Denial Probability numbers, such as 10^{-3} and 10^{-4}, are again only part of the picture. It is essential that even in the absence of called terminal blocking, numerical time-out values be associated with the quoted probabilities. Lacking empirical data, it appears impossible to proceed. The only way out seems to be a hypothetical example. Such an example is offered in Figure 40. This figure depicts only the behavior of the extremely long Access Times. It excludes immediate blocking signals, such as those generated by busy called terminals. Specifically, it is assumed that in the tail of the distribution, the probability density function



Access Denial Probabity = $N_1/(N_1 + N_{00})$

Figure 39. Definition of Access Denial Probability.

Table 28. Proposed Time-Out Values: Three Times the Average Access Time

Applications	Precedence Category			
	I	II	III	IV
Machine I/A Short	0.3	0.6	1.2	2.2
Machine I/A Long	0.6	1.2	1.8	3.0
Human I/A	0.9	1.8	2.7	6.0

Table 29. Proposed Threshold Times and Probabilities of Denial for the Hypothetical Example

Application	Percentile Level	Prob. of Denial	Precedence Category			
			I	II	III	IV
Machine I/A Short	0.999	10^{-3}	0.3*	1.5	4.0	9.0
	0.9999	10^{-4}	0.5	3.5	10.0	25.0
Machine I/A Long	0.999	10^{-3}	0.5	2.5	6.0	12.0
	0.9999	10^{-4}	0.7	5.0	13.0	30.0
Human I/A	0.999	10^{-3}	0.8	4.0	8.0	16.0
	0.9999	10^{-4}	1.0	7.0	16.0	40.0

*All times in seconds.

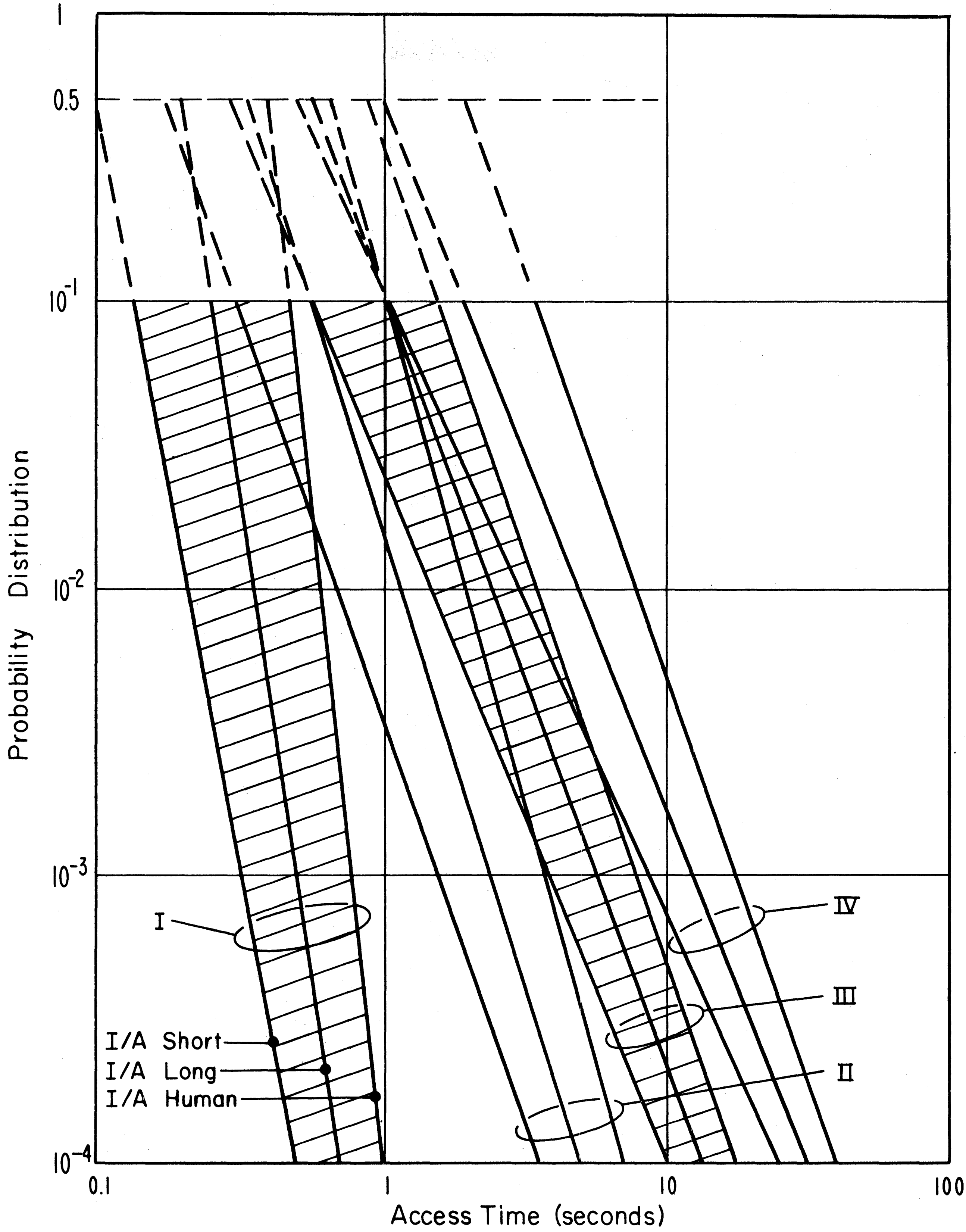


Figure 40. Hypothetical example of Access Time distribution.

(pdf) has the form $cx^{-(1+\alpha)}$, where $c > 0$ and $\alpha > 0$ are constants. It follows that the distribution function is $1 - (c/\alpha)x^{-\alpha}$. The probability that Access Time will exceed a given level x is then $(c/\alpha)x^{-\alpha}$. Both this quantity and the pdf are, of course, linear on the log-log scale. The lines in Figure 40 are determined by two points. One point for each of the 12 lines, may be deduced from the mean and another from the 0.9-percentile values of Table 23. This would resolve the uncertainty at extremes, such as at 0.999 or higher percentiles. If Flash Override is to be nonblocking and to achieve access in no more than 2 seconds (DCA, 1975), then Category I curves should behave as shown in the figure. The 0.99-percentile levels can be assigned to the 1 to 10 second range of the other acceptance categories (DCA, 1975), with successively degraded performances for classes II, III, and IV. The result is the set of the 12 lines shown. Note that the requirements of Figure 40 are more stringent than those of Kelley (1977).

The intersections of the lines with any probability levels, such as 10^{-3} or 10^{-4}, describe fully the Access Denial Probabilities for this model. The numbers are summarized in Table 29. Any user can, if his operations and facilities permit, set his own threshold times. If he were to set the times as suggested in the table, his probabilities of access denial would be as shown. If he were to set them at a different value, such as 2 seconds for all applications and categories, different probabilities would result. These can be deduced from the graphs of Figure 40. A comparison of Tables 28 and 29 reveals that only Precedence Category I is the same for both. All other categories show larger time-outs in Table 29.

Four final comments should be made. First, the numbers of Figure 40 and Table 29 cover a relatively broad range. This makes the worst case choice of the smallest requirement questionable. Second, the individual numbers depend on the choice of the model and thus are inherently uncertain. Third, the realization of actual numbers ultimately depends on the system and its traffic load. Systems can be planned for Access Denial Probability of, say, 10^{-4} at a 2-second time-out, but the proof remains to be shown. Unplanned surges of traffic peaks may render all Access Denial Probability plans unrealistic. And fourth, access blocking caused by busy or inactive distant terminals have not been included above. If this effect dominates the assumed probability levels and renders the service unacceptable, it should be topic of a separate study.

5.4 Transfer Phase

Interim Federal Standard 1033 employs 14 parameters to depict the service during the information transfer phase (see Table 17). As noted further in Table

18, the 14 parameters are divided according to two regimes: performance criteria and transfer functions. There are three performance criteria, called efficiency (or speed) accuracy, and reliability. Likewise, there are three transfer functions, depending on whether one is concerned with bit, block, or message transfers.

In what follows, the approach takes the route indicated in Table 30. The five parameters that deal with bit transfer are treated first, followed by five block transfer parameters. The four message transfer parameters are discussed last. All 14 parameters reflect the character of the high-speed interactive Mode VI. By definition, Mode VI is a binary mode. Its main objective is bit transfer. This section begins with a thorough initial review of bit transfer parameters.

Appropriate groups of successive transactions constitute interactive sessions. An illustration of a successful I/A session is given in Figure 41. The information transfer phase is in the center of the session. It is preceded by the Access Phase (see Section 5.3) and followed by the disengagement phase. Note that the time window of the transfer phase is split into many subintervals. Time is spent sending (transmitting), receiving, processing or reacting to the received information, and waiting in queue for the transmission segments to become available. As will be seen later, the various intervals can be subdivided further according to specific tasks or events. At present, it may suffice to note that for I/A transactions the user process times are expected to be minimal. More time may be spent by segments in queue, even though the virtual circuit is established for the duration of a session. Queues can occur at all PSN's on the path between origination (local) and destination (distant) PSN. If user thinking times become noticeable, they are ordinarily not to be counted against system delivery or transfer times. At issue here is the dependence of service on users' own performance. As noted previously, the ancillary parameters of FS-1033 can include various aspects of this with the help of "user fractions."

5.4.1 Bit Transfer Time

The measurement of Bit Transfer Time begins when the user information bit enters the system through the user-system interface. The system is assumed to be authorized to proceed with its transmission. The measurement ends upon transfer of the corresponding bit from the system to the destination user. Note that the time measurement has meaning only if the bit transfer is successful. Otherwise, the trial is counted as part of the particular failure probability.

The FS-1033 parameter, Bit Transfer Time, is defined as an average of sufficiently many consecutive bit transfer trials. The number of bits involved in the measurement is called a sample. Methods for determining suitable sample sizes

Table 30. The Approach Sequence to the Transfer Phase Number Assignment

Function	Performance Criterion		
	Efficiency of Speed	Accuracy	Reliability
Bit Transfer	1 Para	3 Para	1 Para
Block Transfer	1 Para	3 Para	1 Para
Message Transfer	4 Para		

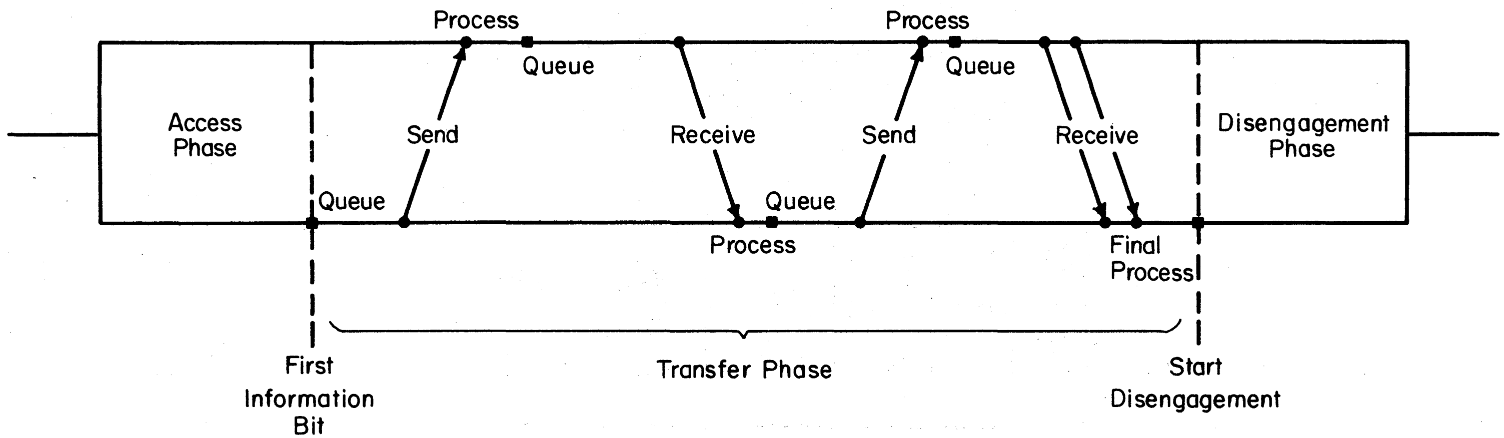


Figure 41. The transfer phase transactions during a successful interactive session.

based on specified Bit Transfer Time values and desired confidence levels have been described by Crow (1979). When constructing a sample, one excludes undesirable bits, such as those belonging to overhead or different application substreams. A sample may extend over several data communications sessions, hours, days, or even months.

After the access phase is successfully completed and an I/A session has begun, the user wants the bit transfer service to take place rapidly. Documentation, such as Martin (1972), DCA (1975), Kelley (1977), and Feldman, et al. (1979), suggests that the average one-way transfer time should not exceed 1 second. The references also indicate tolerable maximum delay values in the 1 to 3 second range for the different I/A traffic categories. Even though a case can be presented for a percentile approach (see Section 5.3), the average or mean value approach is taken here. One reason for this choice is the inescapably large number of bits involved in the estimation of any meaningful bit statistic.

As noted in Figure 41, numerous time elements make up the Bit Transfer Time. Of these, several are the responsibility of the user, while others are system related. Here, one is concerned with delays generated by the system. To describe the system effects the following notation will be helpful. Let

T_A = Arrival or assembly time for the data segment at the initial PSN, plus the equivalent at the destination PSN. At the initial PSN the first bit in may have to wait the longest. The opposite may be true at the destination PSN. If bits arrive serially at 56 Kb/s, then the PSN takes $T_A = 0.0375$ seconds to assemble a 2100-bit segment.

T_X = Transmit or modem time on the packet switched network. It is assumed to be the same on all backbone links, at all PSN's. At the postulated 56 Kb/s rate, a full 2100-bit segment also takes $T_X = 0.0375$ seconds.

T_T = Terrestrial propagation time between two PSN's. For the long distances, perhaps in excess of 1200 km on typical CONUS links, let $T_T = 0.065$ seconds.

T_S = Satellite propagation time for a single up-and-down hop. Assume $T_S = 0.275$ seconds, typically.

T_Q = Average queueing time at a single node. Depending on traffic intensity and system capacities T_Q can vary from zero to large numbers.

T_P = Processing time caused by system and representative of a single PSN delay. Assume that this nodal process time is small, not exceeding 10% of T_X. Thus, $T_P \leq 0.0038$.

T_0 = Output time, needed to format and deliver bits to the destination user.

P = Probability of an error-detection event for a single segment. The error detecting job is performed by the receiving PSN and with the aid of the afore-mentioned 32-bit CRC. In a more detailed treatment one should distinguish between terrestrial and satellite channel bit-error probabilities. For the present, assume channel error probabilities on both types of links to be around 10^{-5} (Kulkarni, et al., 1979; Kirk and Osterholz, 1976). It follows that $P \cong 0.021$ applies to both terrestrial and satellite links.

L_T = Number of terrestrial PSN-to-PSN links on a given path. Usually, L_T = 1, 2, or 3.

L_S = Number of satellite links on a given path. Usually, L_S is either zero or one.

N = ARQ parameter. It denotes the number of repeated segments per each initial detected segment error. For terrestrial and satellite links the appropriate subscript, T or S, is used.

All of the above terms are reasonably clear, except perhaps for the last one. It is explained with the aid of the following three illustrations, Figures 42 to 44. In these figures, as elsewhere, T stands for the total time required by a bit to go from one user, through two or more PSN's, to another user. For simplicity, a single terrestrial path with propagation time T_T is assumed between the two PSN's in all three figures. Figure 42 depicts no detected segment errors. The behavior is typical for all types of ARQ. Note that the Bit Transfer Time, T, consists of a sum of delay elements T_A, T_Q, T_X, T_T, T_P, and T_0. There are no repetitions of elements, because there are no ARQ repeats.

Figure 43 illustrates the presence of a single detected error in segment 5. It also assumes the selected reject type of ARQ, where only the erroneous segment is repeated (Benice and Frey, 1964 a and b; Sastry, 1975). Here, N=1, and for the appropriate bits, the Bit Transfer Time, T, is now increased by exactly the ARQ cycle time length. By the way, this selective ARQ is also called the ideal continuous repeat scheme, because all other schemes result in either larger delays, less throughput, or both. One should also note that segments not directly involved in the cycle, such as segment number 11, may suffer augmented time in the queue, T_Q, because of the ARQ cycle occurrence just ahead.

Figure 44 departs from the ideal ARQ illustration in Figure 43 in two ways. First, it shows three detected segment errors. And second, it postulates a different kind of ARQ scheme, the so-called go-back-N ARQ (Benice and Frey, 1964 a and b; Sastry, 1975; Morris, 1978; Easton, 1980; Lin and Yu, 1980). The go-back-N scheme

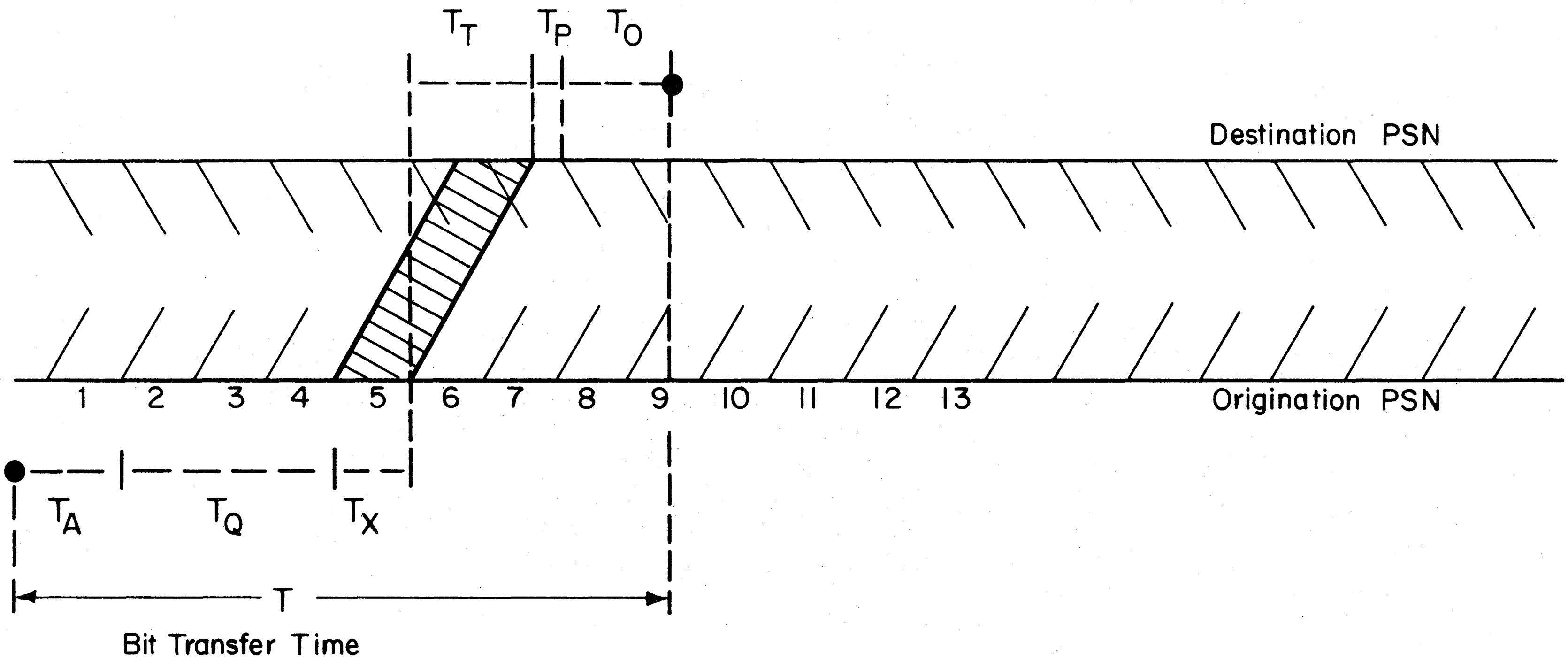


Figure 42. Successful transmission with no ARQ action.

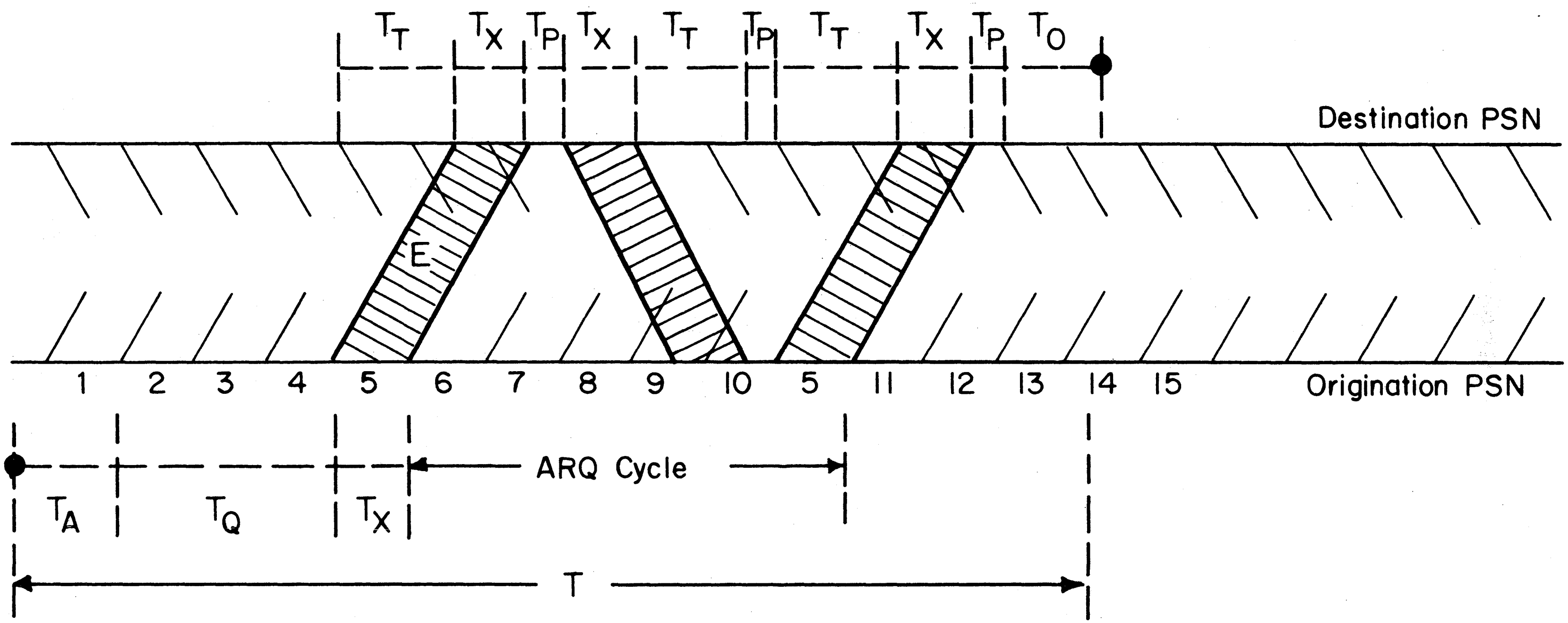


Figure 43. Successful transmission with single detected segment error and a single ARQ cycle on selective reject ARQ.

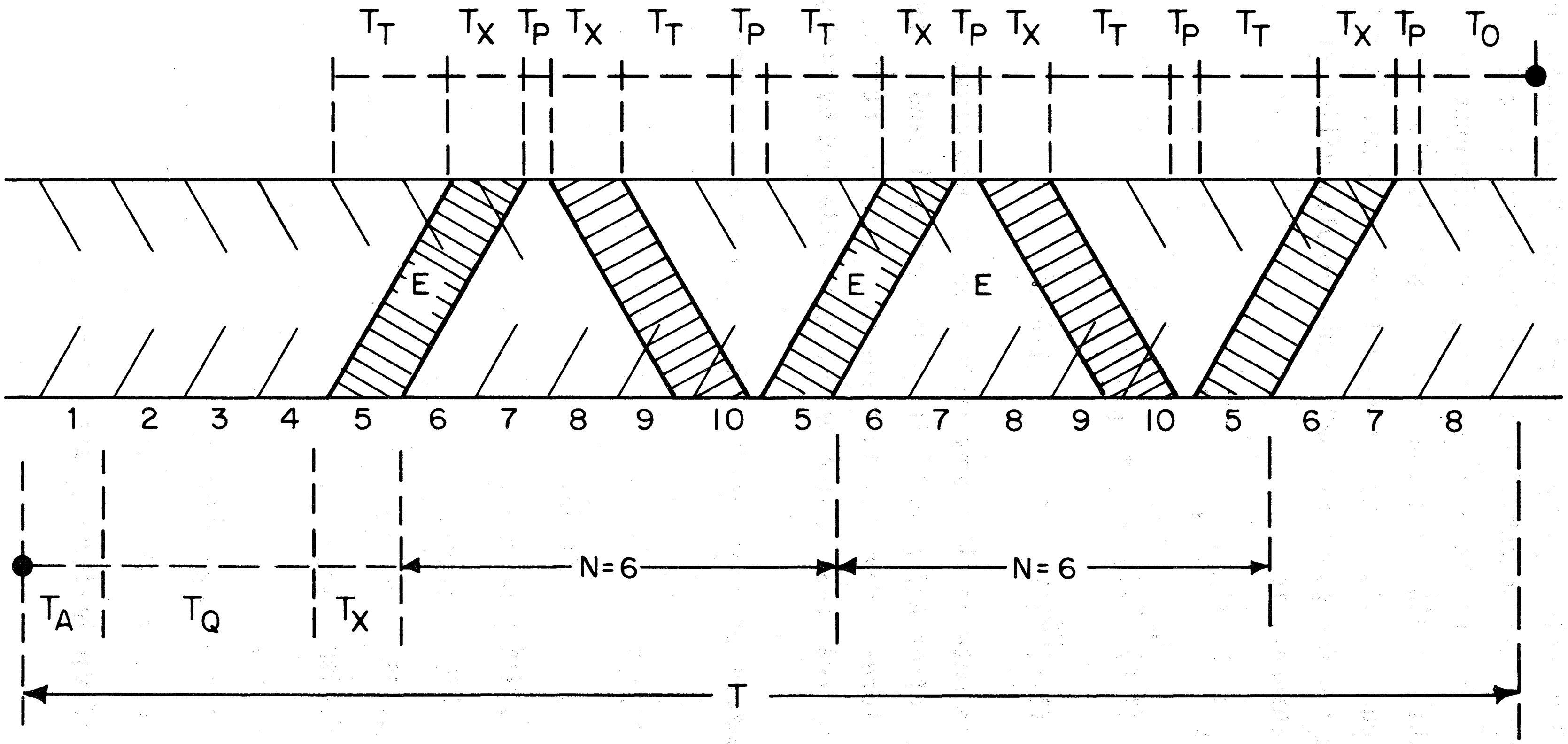


Figure 44. Successful transmission with three detected errors and two ARQ cycles on go-back-N type of ARQ.

is simple and often used. It is, however, more wasteful of time and throughput than the selective ARQ. This is particularly true for systems where the number of segments, N, per ARQ cycle is large. By going back N segments, the scheme in Figure 44 discards up to N-1 potentially correct segments per cycle. Since the ideal scheme discards only the segment in error, the difference is at least qualitatively clear for the two ARQ. The quantitative Bit Transfer Time expressions for the two schemes can be written as a single formula, if one lets N = 1 for the selective ARQ; and N = N for go-back-N ARQ. The formula, valid for an error-free or perfect feedback channel, is

$$\begin{aligned} T = {} & T_A + T_O + (L_T + L_S)T_Q \\ & + L_T[2 - (1-P)^{N_T}](T_X + T_T + T_P)/(1-P)^{N_T} \\ & + L_S[2 - (1-P)^{N_S}](T_X + T_S + T_P)/(1-P)^{N_S} \ . \end{aligned}$$

The main point to note is that many factors enter into computation of T. There are, of course, other ARQ possibilities and they all tend to have different Bit Transfer Time features. However, in cases of practical interest their behavior tends to be bounded by the ideal system on the good side and by the go-back-N on the poorer side.

Based on the numbers assumed earlier, the T values can be computed. Specifically, let

$T_A + T_O = 0.0375$ seconds,

$T_Q = 0$ seconds,

$T_X = 0.0375$ seconds,

$T_T = 0.065$ seconds,

$T_S = 0.275$ seconds,

$T_P = 0.0038$ seconds,

$P = 0.021$,

$N = 1$ for ideal ARQ,

$N_T = 6$ for go-back-N, terrestrial,

$N_S = 17$ for go-back-N, satellite.

Table 31 is the result. As constructed, the numerical values of this table are entirely system oriented. The values are optimal in the sense that they postulate vanishing queueing delays, i.e., $T_Q = 0$. Besides that, they constitute rather basic constraints that the system cannot exceed. The service should not be expected to ask for impossible numbers either. One sees, for instance, that the difference between the two ARQ cases considered is pronounced only for tandem satellite links (Reed and Smetanka, 1977). For all satellite arrangements, the go-back-N Bit Transfer Time is almost twice that of the ideal ARQ. If only terrestrial links are involved, the choice of the ARQ appears to be not that crucial at all. The Bit Transfer Time does, however, grow linearly with the number of tandem links per path.

If the service users were to insist on no more than 0.5 second Bit Transfer Time (a value that appears consistent with the Access Times discussed in Section 5.3.1), then the system routing and error control options would be limited. Only the options indicated in Table 32 would have the potential to deliver Bit Transfer Time lower than 0.5 seconds. At most four terrestrial tandem links may be permitted in the absence of satellite hops. If a satellite hop is allowed, it should not be followed by more than a single terrestrial link. Care should be exercised to control the queueing delay T_Q. If T_Q exceeds 0.1 second, then both Tables 31 and 32 should be modified to reflect that fact. Queue pdf assessment and the associated statistical treatment of delays entails formidable analysis (Kleinrock, 1975 and 1976). Such analysis is beyond the scope of this study. The difficulty is further compounded by the fact that queueing depends on user behavior, traffic generation by users, and various interactions between the system and the users.

The distances, i.e., the number of tandem links, between initial and final PSN pairs are projected to be small integers for the AUTODIN II backbone network. A listing of the planned shortest distances between all $\binom{8}{2} = 28$ PSN pairs is given in Table 33. Note that 16 pairs have the shortest distance of 1, ten have shortest distance of 2, and only two have shortest distance 3. There are no PSN pairs that under normal conditions would require 4 or more links. Table 33 assumes no restrictions on link tandem arrangements for AUTODIN II.

A very rough network-wide average can be estimated by simply computing the mean over the above 1, 2, and 3 link terrestrial paths. For simplicity, let the traffic intensity, be the same over all PSN pairs. Then the Bit Transfer Time average value turns out to be less than 0.2 seconds over the entire terrestrial configuration.

Table 31. System Oriented Bit Transfer Times for Two Kinds of ARQ and Several Terrestrial and Satellite Tandems

Number of Tandem Links		T (Seconds)	
L_T (TER)	L_S (SAT)	Ideal ARQ	Go-Back-N
1	0	0.15*	0.17
0	1	0.37	0.63
2	0	0.26	0.31
1	1	0.48	0.76
0	2	0.70	1.22
3	0	0.37	0.44
2	1	0.59	0.90
1	2	0.81	1.35
0	3	1.03	1.81
4	0	0.48	0.59

*All times in seconds.

Table 32. System Options that may Meet the 0.5-Second Objective for the Bit Transfer Time

Tandem Links		Available ARQ Options
Terrestrial	Satellite	
1	0	Both
2	0	Both
3	0	Both
4	0	Ideal Only
0	1	Ideal Only
1	1	Ideal Only

Table 33. Shortest Distances Between AUTODIN II PSN Pairs

PSN Pairs			No. of Path of Distance			
			1	2	3	4
Albany	-	Andrews	1	1	...	...
Albany	-	Ft. Detrick	---	---	3	6
Albany	-	Gentile	---	2	5	...
Albany	-	Hancock	---	---	3	8
Albany	-	McClellan	---	1	5	...
Albany	-	Norton	1	1	...	...
Albany	-	Tinker	---	2	5	...
Andrews	-	Ft. Detrick	---	1	2	...
Andrews	-	Gentile	1	2	...	...
Andrews	-	Hancock	---	1	3	...
Andrews	-	McClellan	---	3	7	...
Andrews	-	Norton	1	3	...	...
Andrews	-	Tinker	1	2	...	...
Ft. Detrick	-	Gentile	---	3	4	...
Ft. Detrick	-	Hancock	1	1	...	...
Ft. Detrick	-	McClellan	1	2	...	...
Ft. Detrick	-	Norton	---	2	5	...
Ft. Detrick	-	Tinker	1	1	...	...
Gentile	-	Hancock	1	1	...	...
Gentile	-	McClellan	1	3	...	...
Gentile	-	Norton	1	3	...	...
Gentile	-	Tinker	1	3	...	...
Hancock	-	McClellan	1	2	...	...
Hancock	-	Norton	---	2	5	...
Hancock	-	Tinker	---	3	7	...
McClellan	-	Norton	1	2	...	...
McClellan	-	Tinker	1	3	...	...
Norton	-	Tinker	1	3	...	...

Given the system-oriented framework of Tables 31 and 32, as well as the less-than-1-second goal mentioned earlier (DCA, 1975), it appears feasible to propose 0.5 s as a systemwide target for the Bit Transfer Time. The 0.5-second value is to be an average value, intended for all 56 Kb/s, Mode VI, I/A categories. The value is to be interpreted in the sense of the user oriented FS-1033, Bit Transfer Time definition. It applies during the interactive high speed sessions only.

5.4.2 Bit Error Probability

A bit error occurs when the bit in question is transferred from the source user to the intended destination user, but is in error. The correctness or incorrectness of individual bits is observed at the appropriate user-system interface.

Bit Error Probability is the appropriate limit ratio of the total incorrect bits to so obtained total delivered bits. Rather commonly, this Bit Error Probability is called user-to-user or end-to-end Binary Error Rate (BER). Measurement methods for BER and their statistical significance have been developed (Crow and Miles, 1977; Shanmugam and Balaban, 1979). The main feature of all user oriented BER goals is that sometimes they are likely to be met and sometimes not. This fact is illustrated in Table 34. The rows of the table are time series of Bit Error Probability estimates over a particular user data substream. Here row i could be a particular user-to-user pair or a user subcategory, a grouping that would reduce the number of rows involved. The column index j denotes a time interval (e.g., hour, day, operating condition, and so forth) during which the measurement of BER_{ij} is performed. The test time intervals need not be the same for all user pairs or traffic classes. The number of classes could very well be identical to the 12 groupings proposed in Figure 32.

It is well known that the BER_{ij} values can vary considerably, both over time intervals and user categories. For example, on the Dataphone Digital Service (DDS) network there are many hours of error-free operation. There are also times, particularly at certain locations, when the error rates are quite high. Threshold values are apt to be exceeded a certain percentage of the time. The situation here is quite analogous to the percentiles for the Access Time parameter (see Section 5.3.1). For example, a user oriented requirement for BER could quite reasonably be that BER_{ij} observation, perhaps on an hourly basis, should not exceed 10^{-8} more than 1% of the time. Then BER = 10^{-8} would be a 0.99-percentile to be met by the eventual service. Several reports (Kirk and Osterholz, 1976; Kelley, 1977) support the claim that the minimum allowable error threshold will have to be improved, perhaps significantly, over such 0.99-percentile levels as BER = 10^{-5} or 10^{-6}.

Table 34. Variation of Binary Error Rate (BER) Estimates

User Pairs or I/A User Subcategories	Test Time Intervals: 1	2	...	j	...
1	BER_{11}	BER_{12}	...	BER_{1j}	...
2	BER_{21}	BER_{22}	...	BER_{2j}	...
⋮	...	...	...	...	...
i	BER_{i1}	BER_{i2}	...	BER_{ij}	...
⋮	...	...	...	...	...

Because of their user significance, the distributions and percentiles of Bit Error Probability estimates should be investigated in the future. At the present, because of insufficient user oriented service data, the percentile approach appears fruitless. It is abandoned in favor of what may be called the average, or mean value, approach.

Literature suggests typical (i.e., mean or average) user oriented values for BER. Thus, DCA (1975) states that the user-to-user undetected bit error rate through the packet switched network, for subscribers having error controlled access circuits, shall be 10^{-12} or less. For other subscribers without error control on the access lines, the BER value depends largely on the performance of the access lines. A BER goal anywhere between 10^{-4} and 10^{-6} appears realistic here. In addition to this rather broad dispersal of BER objectives, a more basic issue arises here. That is the question whether the accuracy requirements should be determined by the error control implementation on access lines, or vice versa. The point emphasized in FS-1033 is that the actual user needs, whatever their level, should determine the system design.

To gain a feeling for the 10^{-12} and 10^{-5} BER magnitudes, one can turn to two simple examples. Consider two transmission channels. Let the first one have a 56 Kb/s throughput rate. Let the second one have a less efficient throughput, at, say, 10 Kb/s. Then, a 10^{-12} Bit Error Probability corresponds very roughly to a single error per year in the first case, and a single error per more than 5 years, in the second case. On the other hand, 10^{-5} is tantamount to one error every 2 and 10 seconds, respectively.

The 10^{-12} user requirement appears too small to be realistic. Consider the following arguments:

(1) As noted, the 10^{-12} BER level amounts to approximately a single error per year at the fastest throughput rate. Since interactive traffic occurs in short, perhaps infrequent, bursts, this level may be unverifiable in practice by any individual user or user category (Crow and Miles, 1977).

(2) Given an approximate total switch throughput capacity of 10^{7} b/s, over the entire packet switched AUTODIN II network (DCA, 1975), the 10^{-12} BER level would correspond to a single error roughly every 3 hours, somewhere in the total network. From the systems point, it is not certain at all how such rare events could be measured by AUTODIN II system engineers.

(3) It remains to be shown in practice whether the 32-bit CRC error control can deliver anything around 10^{-12} with existing link, node, terminal-hardware and software-technology.

(4) To our knowledge, nobody has shown convincingly that future military users, even in the most critical acceptance categories, will actually require the BER levels to be as low as 10^{-12}.

For these reasons, a more realistic number of Bit Error Probability $\leq 10^{-10}$ is suggested. This is intended initially to be a mean or average (not percentile) value, applicable to all installations with error control on their PSN access lines. As noted in Section 5.3.2, the 10^{-10} goal may be realizable with the planned 32-bit CRC field. The goal appears conservative and measurable.

For terminal-to-PSN lines without error control, the choices are obviously limited. The user will not receive a service performance beyond what is possible with existing technology for line conditioning and equalization. If the lines are capable of 10^{-6} Bit Error Probability, that is precisely what the user will perceive. If the user cannot tolerate so high an error rate, his recourse is to seek other system implementations, perhaps with local error control. Present error control technology can help in many ways. There exist a large number of error-detecting and error-correcting codes (Peterson and Weldon, 1972; Berlekamp, 1968). The codes can be applied in numerous ways, such as forward acting or with feedback. Their potential uses and benefits to AUTODIN II local users should be determined and exploited.

The main issues in the choice between 10^{-12} and 10^{-10} Bit Error Probability user oriented values are presented in Table 35.

5.4.3 Bit Misdelivery Probability

Like Bit Error Probability, Bit Misdelivery Probability pertains to accuracy of bit transfer (see Table 18). The bit misdelivery event occurs in two ways, depending on whether the intended destination does not or does receive its copy. In the first case, the misdelivered bits appear initially as extra bits to the unintended user or users. In the second case, the misdelivered bits appear as correct (successful) to the intended destination, and as extra bits to the unintended destination user(s).

According to FS-1033, bit misdelivery event is a physically measurable (i.e., countable) event. When presented as part of Bit Misdelivery Probability, it answers the question: Of all bits actually sent from A to B, how many were intended for some destination other than B?

Bit Misdelivery Probability is defined as the ratio of recognized misdelivered bits to terminal B, to total intended bits for terminal B. When measuring the service values of Bit Misdelivery Probability, the necessary statistical methods and confidence limits must be employed (Crow and Miles, 1977).

Table 35. The Main Issues for Choosing the User Oriented Bit Error Probability Number

Issue	Candidate Bit Error Probability 10^{-12}	Candidate Bit Error Probability 10^{-10}
Frequency of Errors on the Output of a Single 56 Kb/s Link	$\frac{1}{\text{Year}}$	$\frac{1}{\text{3 Days}}$
Networkwide Frequency of Output Errors	$\frac{1}{\text{3 Hours}}$	$\frac{1}{\text{2 Min.}}$
Can the BER Level be Achieved with Planned 32-Bit CRC?	Not Sure	Yes

AUTODIN II Performance Specification (DCA, 1975) calls for probability of segment misdelivery of less than 10^{-11}. Since the ratio of misdelivered segments is synonynous with the same ratio of misdelivered bits, the above requirement calls for Bit Misdelivery Probability of 10^{-11}.

Returning to Section 5.3.2, one recalls that the value for Incorrect Access Probability was considered in the 10^{-10} to 10^{-12} range. The fact that the Misdelivery probability value of 10^{-11} falls within the range appears reasonable from the system and user points of view.

5.4.4 Extra Bit Probability

Extra Bit Probability is the ratio of total extra bits received to total received bits. An extra bit outcome occurs when a bit transferred to the destination user turns out to be not from the source user currently engaged in the information transfer process. Unless misdelivered bits are differentiated from extra bits, both will appear as extra bits.

The effective harm or cost caused by extra bits to the Mode VI user appears to be no less than that of bit errors. Extra bits could result in serious consequences both for system security elements and for the user's own security devices. Therefore (see Section 5.4.2), a candidate value for the selected digital I/A mode should also be in the 10^{-10} to 10^{-12} range. Further study of AUTODIN II user vulnerability to extra bits should be undertaken to pin down the issues and values with more confidence (see Figure 45).

5.4.5 Bit Loss Probability

Bit Loss Probability is defined as the ratio of total lost bits to total transmitted bits. A lost bit event occurs when a bit transferred from the source user, through its user-system interface, fails to pass through the user-system interface (i.e., becomes lost) at the destination user end. It has a direct impact on the often used parameter "bit count integrity" (see Tables 15 and 19).

The effective harm or cost caused by bit loss to the military Mode VI user appears to be comparable to those of bit errors and extra bits (see Sections 5.4.2 and 5.4.4). Communication security devices on both sides of the system-user interface would be affected, as noted earlier. The proposed value for Bit Loss Probability should accordingly be somewhere between 10^{-10} and 10^{-12}, perhaps 10^{-11}. More detailed study is needed to identify bit loss effects on various acceptance and COMSEC application classes (Figure 45).

5.4.6 Block Transfer Time

Since Mode VI is a binary mode, the interactive users may arrange their information into formats of their own. Thus, for their own use the host systems may

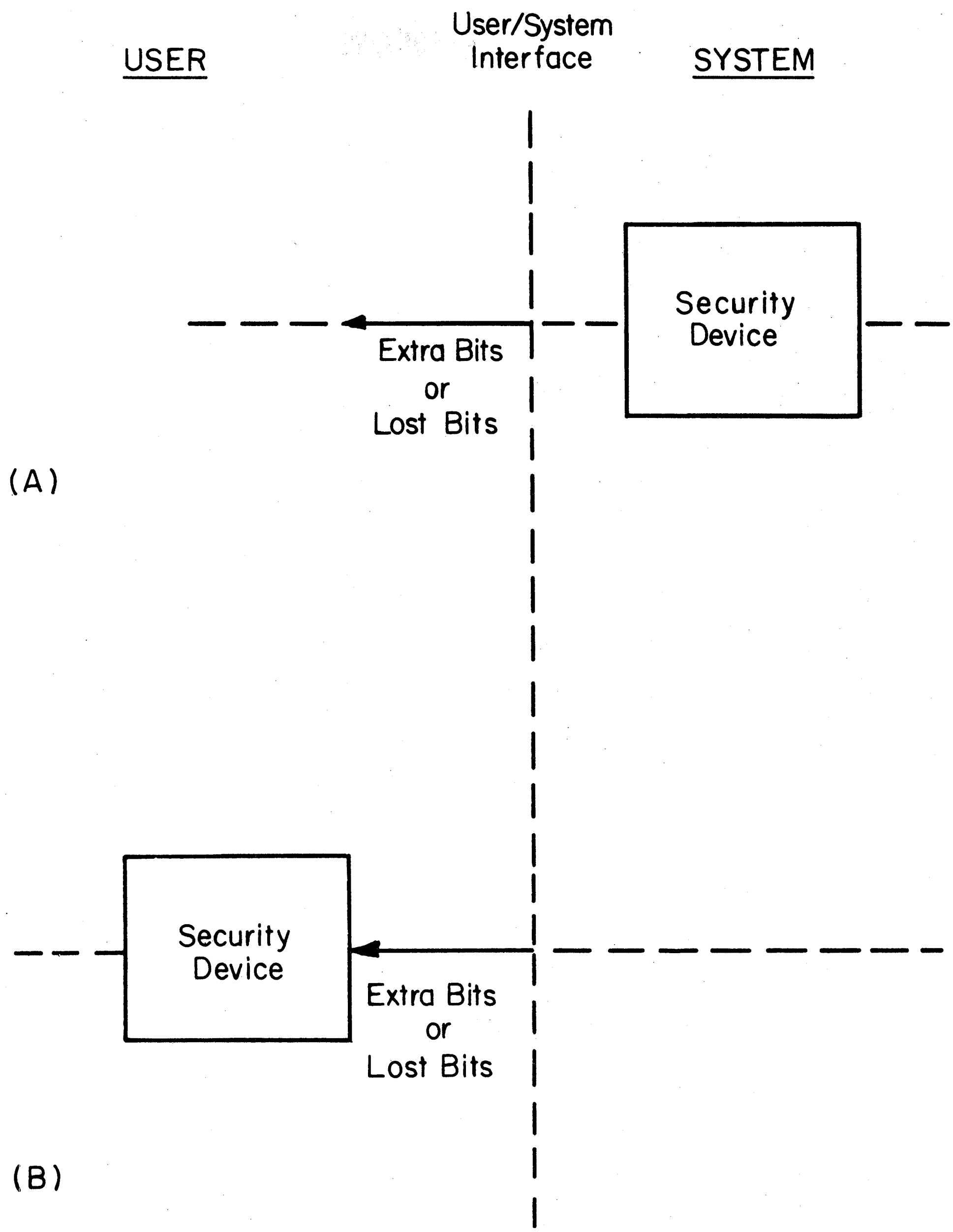


Figure 45. The extra or lost bits and COMSEC devices at the user-system interface.

define host unique characters, words, frames, card images, or any blocks, in general. Such blocks consist strictly of contiguous user information bits. They do not contain system overhead and parity bits. Thus, the blocks of FS-1033 differ from the conventional uses elsewhere (ANSI, 1971; ANSI, 1980). If blocks are not longer than the 1952-bit text field of a segment (see Figure 28), then the variable length of the text allows the blocks to be transmitted as a whole.

Just as for bit transfer, the block transfer function commences when the entire user information block has been sent through the user-system interface and the system has been told to proceed with transmission. The block transfer function ends when the entire block has been successfully transferred from the system, through the user-system interface, to the destination user.

Block Transfer Time is the elapsed time between the beginning and end of the block transfer function. In this and all other respects, the Block Transfer Time definition is comparable to the definition for Bit Transfer Time, T (Section 5.4.1). The segment assembly time, T_A, is needed for short blocks at the first PSN, but there is negligible output time, T_0, at the last PSN. For long blocks, namely those that occupy the whole segment, no T_A is involved at the first PSN. However, delay is caused by the dispatch of the block over local lines to the destination user. Then T_0 cannot be ignored. Accordingly, the previous T formula and computed T values (see Table 31) apply for blocks as well as bits.

As for Bit Transfer Time, it is proposed that the Block Transfer Time shall be 0.5 seconds. This 0.5 second figure is to be considered as an average user oriented value over all Mode VI, 56 Kb/s service, I/A subcategories. The maximum tolerable value for the Block Transfer Time may depend on application and precedence class. As stated by DCA (1975), the permissible maximum values for I/A traffic may range from 1 to 3 seconds.

The following critical question must be asked. Why should the 0.5 second requirement apply to both Block and Bit Transfer Times, without any discrimination for the transaction type (machine, human) or its priority class (Flash Override, Priority, etc.)? Special users, after all, require special transfer times, most likely in the 0.3 to 1 second range (DCA, 1975). Here one has taken the apparently easy way out, because of the invariant physical facts involved in the formula for T (see Section 5.4.1). The propagation times T_T and T_S are determined basically by the speed of light and related constraints. The transmit times T_A, T_X, and T_0 can be reduced by departing from the 56 Kb/s rate assumed for the selected mode. However, their effect would not be major. Likewise, only relatively minor effects can be generated by lowering the segment error rates P_T and P_S. In the systems sense, the tuning of P's may be too costly and difficult to do. What remains are the numbers, L_T and L_S, for terrestrial and satellite links per path. As seen from Tables 31 and 32, the exclusion of satellite hops (i.e., setting $L_S = 0$) has a major impact on keeping T low. Up to 3 tandem terrestrial links, and no satellite hops, yield Block Transfer Time values shorter than 0.5 seconds for the ordinary go-back-N repeat scheme. If this issue is important to AUTODIN II users, it should be addressed in more detail (see Feldman, et al., 1979).

5.4.7 Block Error Probability

Block Error Probability is defined as the ratio of total incorrect blocks to total delivered blocks. As for Bit Error Probability, delivery occurs through the user-system interface at the destination user site. A block is said to be in error if it contains one or more bit errors.

Extreme contrived cases help relate Block Error Probability to Bit Error Probability (Figure 46). If every bit of every erroneous block were in error, see Part (A), then the Block and Bit Error Probabilities would be indistinguishable. On the other hand, if exactly one bit of every erroneous n-bit block were in error, as in Part (B), then Block Error Probability would be n times the Bit Error Probability. The true relationship, Part (C), is usually found somewhere in the middle between 1 and n. When the block length n is small, such as $n \leq 10$ for short machine I/A messages, the difference between the two probabilities is not significant. When n becomes large, this writer prefers a "rule of the thumb" value of $\sqrt{n}$.

From the user point of view, block errors manifest as groupings of information bit errors in the received text. By the above mentioned "rule of the thumb", one block error may generate roughly $\sqrt{n}$ bit errors. For a n = 1952 data field of a segment, $\sqrt{n} \cong 44$ bit errors may result from each segment error. This, of course, is the case only for the maximum length of a continuously transmitted block.

It is recommended that, on the average, the Block Error Probability for n-bit blocks should not exceed $\sqrt{n}\ 10^{-10}$ for all interactive, 56 Kb/s, Mode VI traffic. This block accuracy requirement is equivalent to the 10^{-10} bit performance level summarized in Table 35. It appears realistic for the planned 32-bit CRC error control.

The main features of the proposed user oriented Block Error Probability requirement are shown in Figure 47. The requirement, as shown, is the same for all priority classes and all applications of I/A, 56 Kb/s, Mode VI transactions. If a single, mean or average, number must be chosen for the user oriented Block Error Probability, it should be in the neighborhood of 10^{-9}.

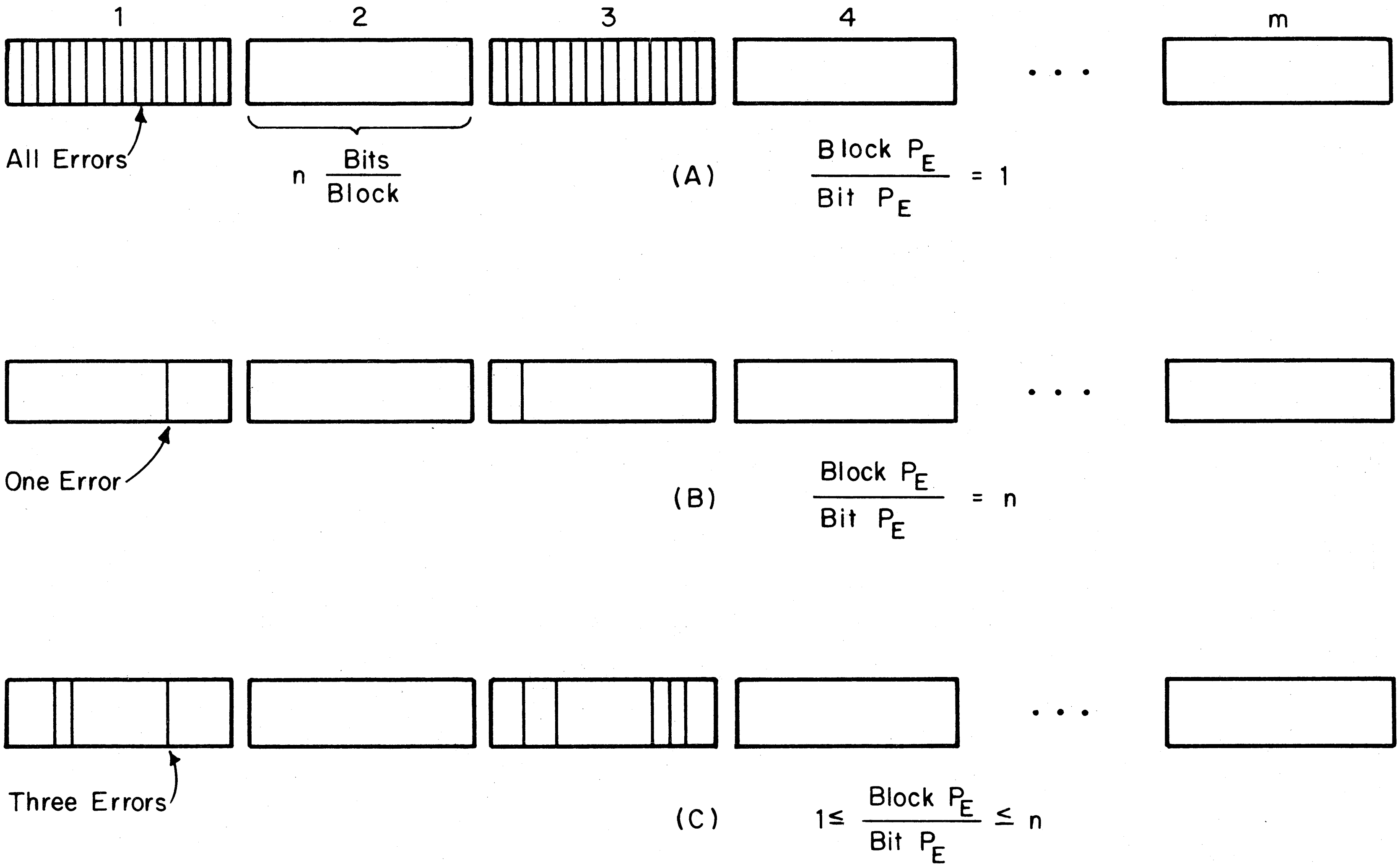


Figure 46. Extreme relationships between bit and block error probabilities.

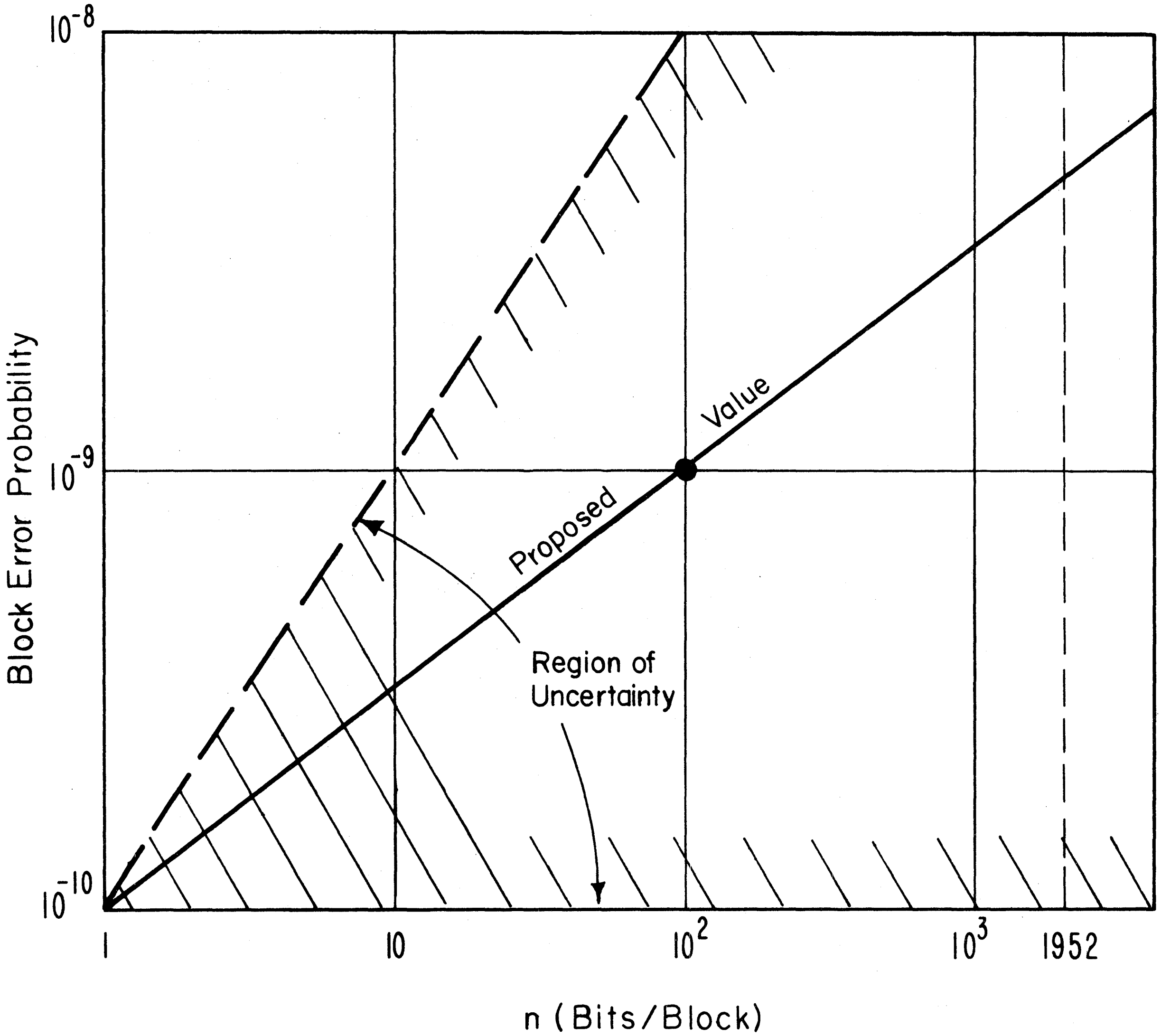


Figure 47. Proposed Block Error Probability value as function of block length.

5.4.8 Block Misdelivery Probability

Block Misdelivery Probability is the ratio of total misdelivered blocks to total transferred blocks. The definition is nearly identical to that of Bit Misdelivery Probability in Section 5.4.3, except that an n-bit entity, the block, has replaced the single bit.

It is recommended that the Block Misdelivery Probability target value be around 10^{-11} for all block lengths n. This applies for the total assumed I/A traffic mode, under the usual statistical averaging constraints. The number is supported by DCA (1975) documentation, the user oriented Bit Misdelivery Probability objective of Section 5.4.3, as well as the practical contention that the severity of block misdelivery to a user may be higher than from a block error event. The latter event probability straddles the 10^{-9} value, as seen earlier in Figure 47.

5.4.9 Extra Block Probability

Extra Block Probability is defined as the ratio of total extra blocks received to total received blocks. The specifics of FS-1033 definition for extra blocks are again nearly indistinguishable from extra bits definition (see Section 5.4.4).

As a single candidate number for user oriented Extra Block Probability, 10^{-10} is recommended. It is roughly consistent with the 10^{-10} to 10^{-12} range recommended for Extra Bit Probability in Section 5.4.4, the 10^{-9} Block Error Probability in Section 5.4.7, as well as the 10^{-11} value suggested for Block Misdelivery Probability in Section 5.4.8. The COMSEC concerns expressed earlier for extra bits apply also to extra blocks (see Figure 45).

5.4.10 Block Loss Probability

Block Loss Probability is defined as the ratio of total lost blocks to total transmitted blocks. The definition is analogous to the Bit Loss Probability.

Based on the numerical values quoted in Sections 5.4.5, 5.4.8, and 5.4.9, a range of 10^{-10} to 10^{-11} appears reasonable for user oriented Block Loss Probability. If a single number is required, it can be chosen in the middle of that range, namely at $3 \cdot 10^{-11}$.

Here, and likewise for the entire set of 5 block transfer parameters, more specific user service numbers may be needed. If so, appropriate user need studies that emphasize communications security implementations seem to be desirable.

5.4.11 Message transfer phase

In the FS-1033 format, see Table 18, message transfer is characterized by 4 primary user-oriented performance parameters, plus one secondary parameter. The primary parameters belong under the performance criterion called Efficiency, and are called:

Bit Transfer Rate,
Block Transfer Rate,
Bit Rate Efficiency,
Block Rate Efficiency.

In what follows, these four parameters are treated as a group, because their definitions, issues and value choices are closely related. The Federal Standard uses the term block to denote any contiguous set of bits, where the number of bits can be larger than or equal to unity. If one focuses on the case $n = 1$, the terms bit and block become interchangable. In this section, the term block refers to the ordinary data block when $n > 1$, or to a single binary digit when $n = 1$.

Subject to the above convention, only two primary message transfer phase parameters remain to be distinguished. They are the Block Transfer Rate and the Block Rate Efficiency.

Block Transfer Rate is defined as the number of successful block transfers during a performance measurement period, divided by the duration of that period. The dimensions of Block Transfer Rate are blocks per second.

The Block Rate Efficiency is defined as the ratio

$$\frac{\text{(Block Transfer Rate)(Average Block Length)}}{\text{Signaling Rate of the Communication Service}} .$$

The new term here is the "signaling" rate. In FS-1033, it is defined as the maximum rate in bits per second, at which binary information could be transferred unidirectionally between users under conditions of:

(1) Continuous uninterrupted transmission.

(2) No overhead information (namely, no control, header, repeat, or redundancy of any kind).

This signaling terminology of FS-1033 should not be confused with the terms employed in telephony. In the latter case, signaling refers to overhead information used for various circuit switched network controls.

The average block length is, of course, unity for Bit Transfer Rate.

In the interactive rapid exchanges during an I/A session, the throughput rates and their efficiencies may or may not have major significance to the individual user (Brayer, 1978, a and b). Rates appear to be relatively insignificant to the short message category. However, they may be significant factors to those users who send and receive long I/A messages, such as bulk messages with more than 100,000 bits per message. Even then, a typical user may be hard pressed to quote what minimum rate he requires. Pertinent user oriented data appear to be missing.

From the systems point of view, the vehicle for throughput is the same ARQ framework discussed earlier in Section 5.4.1. Of particular importance are the illustrations of that section, notably Figures 42 to 44. Assume, as before, that the feedback links do not contribute any errors. Then in the presence of detected forward link errors, the two ARQ schemes are seen differently by the receiver. The matter is illustrated in Figure 48. Only a fraction of all arriving segments are acceptable to the receiver. If the probability of detected segment error is P, then for each link

$$\frac{\text{Accepted Segments}}{\text{Total Segments}} = \frac{1 - P}{1 + (N-1)P} .$$

This simple formula is valid for both ARQ schemes discussed here. For go-back-N ARQ one substitutes cycle length N_T or N_S (see page 137) for N, while for the ideal selective ARQ one sets N = 1. As a consequence of this, the expressions for both the Block Transfer Rate and Block Rate Efficiency are readily derived. One obtains according to the definition of Federal Standard 1033:

$$\text{Block Transfer Rate} = \frac{56{,}000 \left\lfloor \frac{1952}{n} \right\rfloor (1 - P)^{L_T + L_S}}{2096\ M[1 + (N_T-1)P]^{L_T}\ [1 + (N_S-1)P]^{L_S}} ,$$

$$\text{Block Rate Efficiency} = \frac{n \left\lfloor \frac{1952}{n} \right\rfloor (1 - P)^{L_T + L_S}}{1952\ [1 + (N_T-1)P]^{L_T}\ [1 + (N_S-1)P]^{L_S}} ,$$

where integer M stands for the average number of simultaneous sessions over the representative packet stream between the respective end users. In effect then, a given user pair is assumed to have access to 1/M of the passing segments. The symbol $\lfloor x \rfloor$ in the above denotes the integer part of real number x. It represents "block division" or quantizing loss that occurs when one tries to pack a field of 1952 bit spaces with blocks of length n.

Both Block Transfer Rate and Block Rate Efficiency formulas are system oriented. They show what certain type systems could deliver to the user, as a performance level, if the systems fit the assumed model. The model includes the ideal and go-back-N ARQ options. It assumes the same segment error probability, P, for both terrestrial and satellite links. Furthermore, it assumes 1952 information bits in a packet of total length 2096. The Block Transfer Rate, as given, does depend on M, the number of multiple user pairs. However, Block Rate Efficiency does not. The model totally ignores queueing effects.

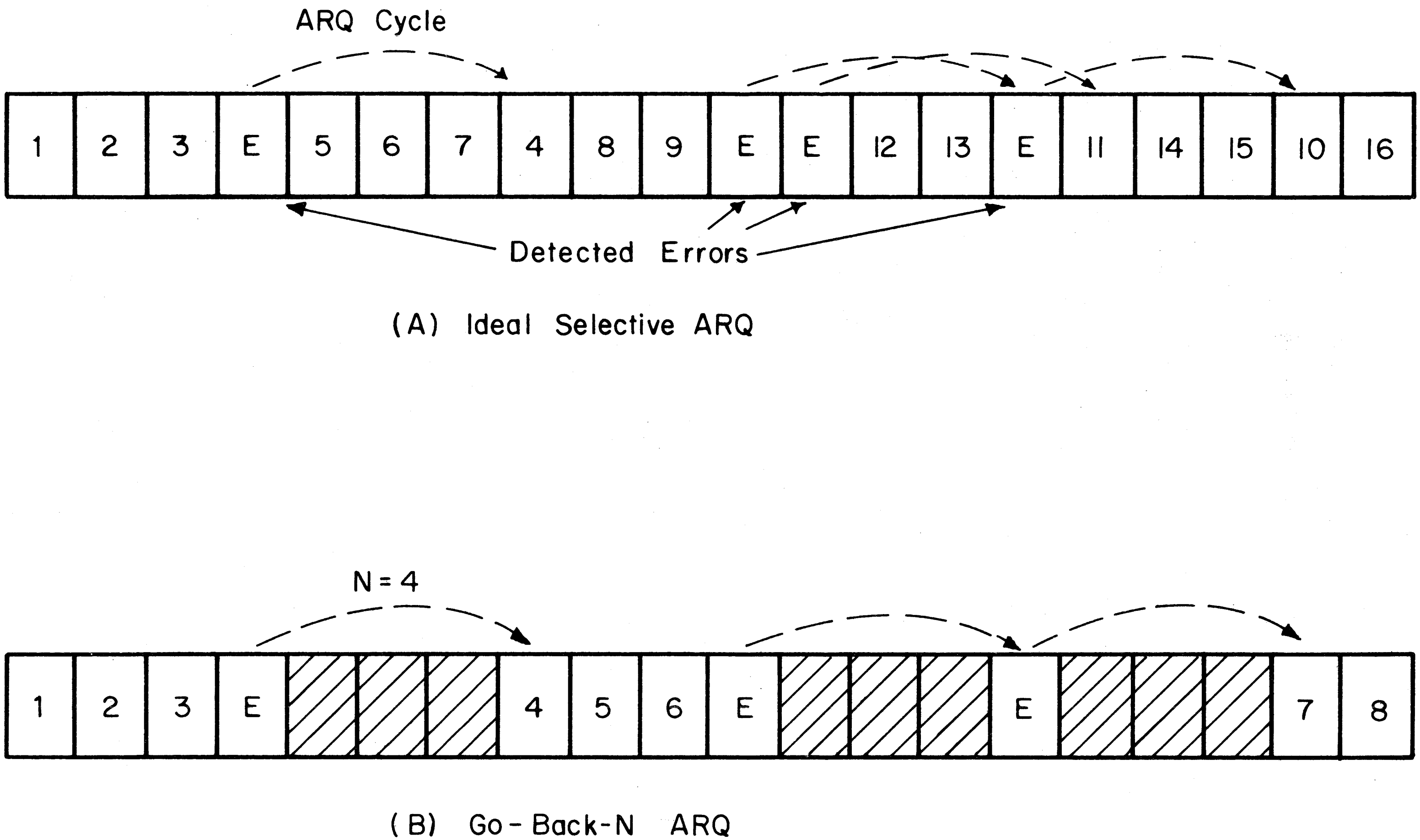


Figure 48. Segment arrivals at the receiver, for both ARQ's and cycle length N=4.

The large number of parameters, namely n, M, P, L_T, L_S, N_T, and N_S, makes a graphic presentation of Block Transfer Rate and Block Rate Efficiency effects cumbersome. To proceed, assume P = 0.021. It corresponds to channel bit error probability of 10^{-5} on both terrestrial and satellite links. Assume a typical value M = 3. Also assume that one is primarily interested in those ARQ system operations that show potential for meeting the 0.5 second Bit Transfer Time objective mentioned in Section 5.4.1. The potential ARQ options are listed above in Table 32.

Table 36 shows Block Transfer Rates estimated accordingly. All the rates given are in blocks per second. Since the block length is n, the column under the heading n = 1 represents the FS-1033 performance parameter, Bit Transfer Rate. Note that the bit content of block transmissions decreases with n, but not at a fast rate. For small n, such as n = 4 or n = 16, the number of bits per second is the same 17,020 as for n = 1. The quantization loss is more manifest for the largest n. The effect of ARQ is relatively more pronounced. Depending on the number of links, the go-back-N ARQ shows a 10% to 20% lower Block Transfer Rate than the ideal ARQ. The choice of the eventual AUTODIN II ARQ scheme (Sastry, 1975; Morris, 1978; Easton, 1980) should take this into consideration.

Table 37 depicts the Block Rate Efficiency computed with the second above formula. The same assumptions are made here as before. Notable among the assumptions made for this model is the absence of queueing losses. If one interprets condition (1) to mean that continuous transmissions, by definition, do not involve any queues, then the absence of queues (i.e., T_Q = 0) appears to be a serious shortcoming of Table 37. Unfortunately, and as mentioned before, planned AUTODIN II queueing statistics were not available for this study. Thus, they could not be incorporated in these tables and other related parameter estimates.

The n = 1 column shows the Block Rate Efficiency for bit oriented transactions. For small n, such as n = 1, 4, 16, there is no quantizing loss and the Block Rate Efficiency values are indistinguishable. There is a nominal efficiency loss for increasing n. As for Block Transfer Rates, the Block Rate Efficiency of the go-back-N ARQ is some 10% to 20% below the ideal ARQ.

The two tables show what certain system models are capable of doing. Clearly, this may differ from what the Mode VI, I/A, 56 Kb/s transaction users need. From the throughput experiences of existing packet networks, such as ARPANET (Kleinrock, Naylor, and, Opderbeck, 1976; Kleinrock and Opderbeck, 1977) and ALOHA (Abramson, 1977) rate efficiences in the 10% to 60% range are common, because of the previously mentioned queueing delays, repetitions and overheads. If a 0.9-percentile level

Table 36. Block Transfer Rate Estimates for Repeat Probability P = 0.021 and M = 3 Simultaneous Sessions

Number of Links		ARQ Type		Number of Bits Per Block n					
L_T	L_S	Ideal	Go-Back-N	1	4	16	64	256	1024
1	0	✓		17,020*	4,260	1,060	262	61	9
1	0		✓	15,400	3,860	960	237	55	8
2	0	✓		16,660	4,160	1,040	256	60	8
2	0		✓	13,940	3,490	870	214	50	7
3	0	✓		16,310	4,080	1,020	251	58	8
3	0		✓	12,610	3,160	780	194	45	6
4	0	✓		15,970	3,990	990	245	57	8
0	1	✓		17,020	4,260	1,060	262	61	9
1	1	✓		16,660	4,160	1,040	256	60	8

*All rates in blocks per second.

Table 37. Block Rate Efficiency Estimates for Repeat Probability P = 0.021

Number of Links		ARQ Type		Number of Bits Per Block n					
L_T	L_S	Ideal	Go-Back-N	1	4	16	64	256	1024
1	0	✓		0.979	0.979	0.979	0.963	0.899	0.514
1	0		✓	0.886	0.886	0.886	0.870	0.814	0.465
2	0	✓		0.958	0.958	0.958	0.942	0.879	0.502
2	0		✓	0.802	0.802	0.802	0.789	0.736	0.421
3	0	✓		0.938	0.938	0.938	0.923	0.861	0.492
3	0		✓	0.726	0.726	0.726	0.714	0.666	0.381
4	0	✓		0.919	0.919	0.919	0.904	0.844	0.482
0	1	✓		0.979	0.979	0.979	0.963	0.899	0.514
1	1	✓		0.958	0.958	0.958	0.870	0.879	0.502

is to be used in the sense that the rate should be above the level, then an efficiency of 0.5 would be a reasonable, though unsubstantiated, service objective. To proceed with user oriented numbers for Block Transfer Rate and Block Rate Efficiency, more data on real life user needs appear necessary.

5.5 Disengagement Phase

Disengagement phase is the last of the three functional phases. See Figure 41 and Table 18. It appears that from the users perspective, the disengagement phase is less important than the access and transfer phases. Unsuccessful disengagement may cause the system merely to ignore or to misinterpret the user's next access attempt.

The disengagement phase is represented by two primary FS-1033 parameters. They are called Disengagement Time and Disengagement Denial Probability.

5.5.1 Disengagement Time

The parameter, Disengagement Time, is defined in FS-1033 as the average value of lapsed time between the start of a disengagement attempt and successful disengagement. The successful disengagement or connection closing event occurs when a disengagement confirmation signal is received by a user at his user-system interface, and within the specified maximum time allowed for disengagement. If certain interactive Mode VI users prefer not to use the disengagement confirmation signal, then the system proceeds to terminate the I/A session upon receipt of a disengagement request. The request, at least in principle, could be originated by either user, or by system time-out condition (i.e., interrupt). The last condition may be the result of nonactivity by either user or both users during the I/A session.

In analogy to the Access Time definition in Section 5.3.1, the average value of a random variable - Disengagement Time - is specified by FS-1033. The average could be the arithmetic mean observed for the time series at a specific user-to-user pair. Or it could be the average over a category of similar users.

Let us assume that a local user submits his disengagement request (i.e., CLOSE order) to the local PSN, and waits for the disengagement confirmation. The key delay elements of such a procedure are nearly the same as illustrated in Figure 31 for Access Time. Of course, the queues should be shorter for disengagement. Likewise, the disengagement confirmation should take less time at the distant user PSN. The transmission and propagation times, however, should be comparable.

In view of the above, one anticipates that the users should have a shorter Disengagement Time than Access Time. A 0.05-second average value seems appropriate for all subcategories of the interactive, 56 Kb/s, Mode VI service. It would be an insignificant time from the user's point of view, and it would not tie up system resources too long after the session work is finished. If one is interested in the 0.9-percentile value for the Disengagement Time, it could plausibly be set around 0.1 second.

5.5.2 Disengagement Denial Probability

Federal Standard 1033 defines Disengagement Denial Probability as the ratio of total disengagement attempts that result in disengagement denial to total disengagement attempts. Disengagement denial is a disengagement event that does not qualify as successful (see Section 5.5.1). The disengagement denial occurrence may be indicated by the absence of the disengagement confirmation signal within the specified maximum time limit. In instances where a nonoriginating user cannot initiate or prevent disengagement, disengagement denial can only be caused by the system. In those system-user arrangements that function without confirmation signals, successful and unsuccessful disengagement confirmations can only be observed by re-accessing the system.

The time-out condition, according to FS-1033, is to be the maximum permitted Disengagement Time, equal to or greater than three times the nominal value of the parameter Disengagement Time. Since the average value for Disengagement Time was set at 0.05 seconds in Section 5.5.1, the corresponding time-out value would be 0.15 seconds. Whether this value is acceptable to military AUTODIN II users remains to be verified.

Since disengagement denials eventually result in automatic time-out, their consequences do not appear to be more serious than those of access denials. This seems to be true for both the users and the system. Accordingly, the maximum number of 10^{-3} is assigned to Disengagement Denial Probability. Combined with the previously mentioned 0.15 second time-out, this user requirement would clear the system considerably faster than the Access Time requirements would fill the system (see Figures 33 to 40).

5.6 Outages

Outages pertain to service availability or lack thereof. Outages are described by the so-called secondary parameters of FS-1033. The word "secondary" is not meant to delegate availability to a lower level of importance for digital services. Rather, it denotes that the parameters are either equal to, derived from, or are strongly correlated with, previously reviewed primary parameters.

5.6.1 Outage definition

Three secondary parameters are listed in Table 18. They pertain to outages and are defined in terms of primary parameters, such as Bit Transfer Rate, Bit

Error Probability, Bit Misdelivery Probability, Bit Loss Probability, and Extra Bit Probability. When the value of any of these primary parameters is worse than its assigned outage threshold value, the service is said to be unavailable and the system is said to be in a potential outage state. With each state one associates a reference time period. The system must be in the potential outage state longer than the reference time period, for the outage to be formally recognized as such.

Figure 49 offers a simplified definition of the outage determination. In this figure, three pertinent parameters are shown to cause outage events. All three parameters have their nominal values and specified outage thresholds. At least one observed value must violate its threshold for more than the given reference time period, for an outage to be declared. As shown, the threshold time period may be the same for all parameters; or it may be custom fitted to parameters and service applications.

Methods for deriving outage threshold values in terms of nominal parameter values are indicated in FS-1033. It is suggested that the outage threshold for user information Bit Transfer Rate be set at 1/3 of the nominal value for that rate. Furthermore, the outage thresholds for the pertinent bit (or block) transfer probabilities are to be defined simply as the square roots of the respective nominal values (see Appendix B and GSA, 1979).

For example, if the nominal 0.9-percentile value for Bit Transfer Rate (see Table 35) were 50% of the maximum possible over a single terrestrial link with ideal selective ARQ, or 8,510 bits per second, then the outage threshold value for Bit Transfer Rate would be 2,836 bits per second. The Bit Error Probability outage threshold would be 10^{-5} for the nominal requirement of 10^{-10} (see Section 5.4.2). Finally, the Bit Misdelivery, Extra Bit, and Bit Loss Probabilities would have outage thresholds of $3(10^{-6})$ (see Sections 5.4.3, 5.4.4, and 5.4.5).

The reference time period is also called the performance measurement period in FS-1033. Its duration must be sufficient to estimate the parameters with reasonable confidence (Crow and Miles, 1977; Crow, 1979). For a nominal Bit Transfer Rate of 8,510 bits per second, the estimation of Bit Error Probability outage levels around 10^{-5} may consume 10 to 25 seconds, for example.

The FS-1033 outage parameters are Service Time Between Outages, Outage Duration, and Outage Probability. Their numerical value assignment is discussed next.

5.6.2 Service Time Between Outages

Service Time Between Outages is denoted as secondary in Table 18. It pertains to the function of service continuation. Service Time Between Outages is defined

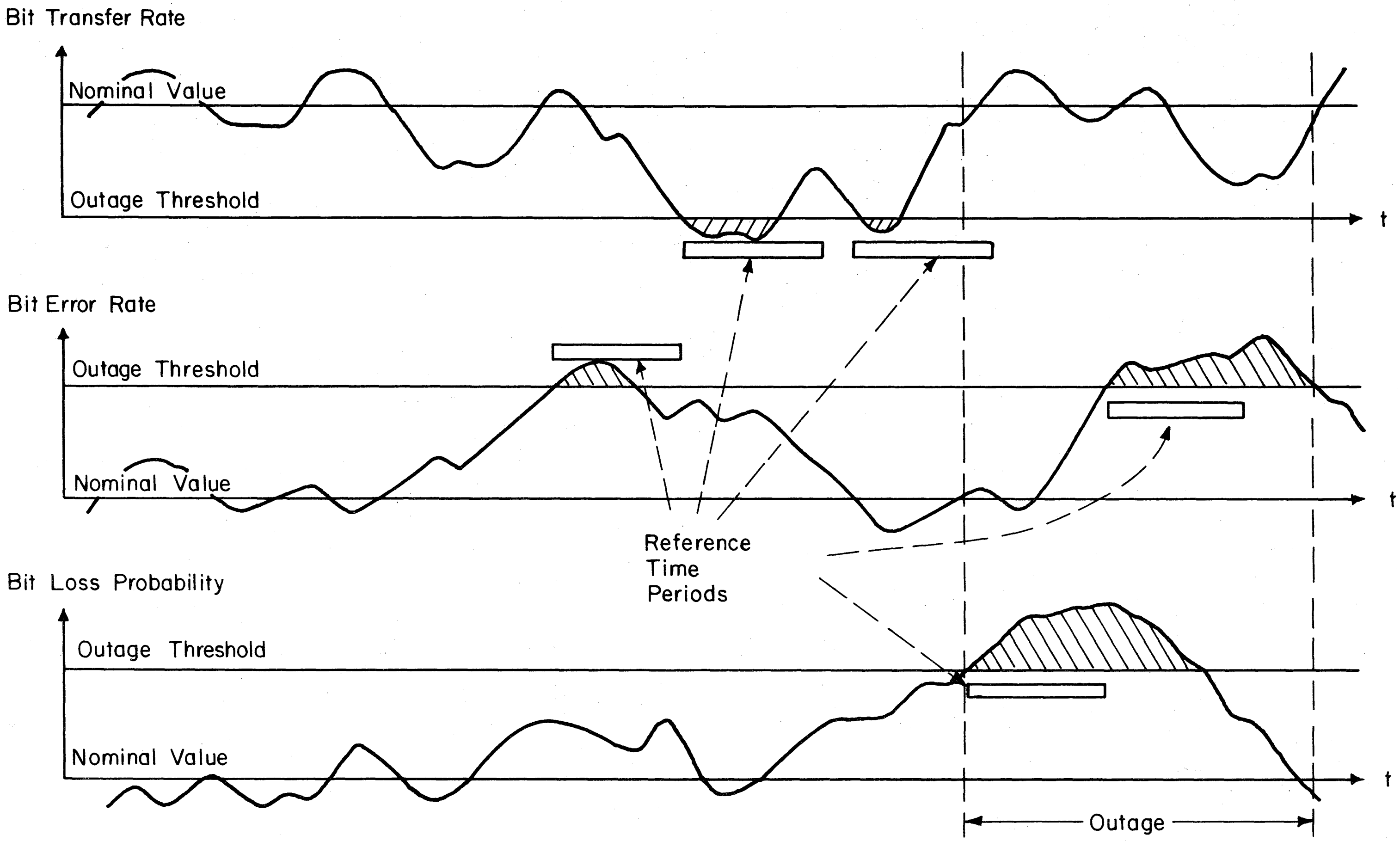


Figure 49. Outages as functions of outage thresholds and reference time periods.

in FS-1033 as the average value of elapsed user information time between outages. Subject to careful statistical definition of start and end epochs (Appendix B), this parameter provides an unbiased estimate of the traditional reliability parameter Mean Time Between Failures (MTBF).

User oriented values for Service Time Between Outages, be they the FS-1033 recommended averages or percentiles, appear difficult to obtain for several reasons. First, the existing and projected user requirements for this parameter are unknown. And second, Service Time Between Outages is somehow related to other secondary and primary user oriented performance parameters. Like availability, Service Time Between Outages appears important to interactive, 56 Kb/s, Mode VI traffic users. It should be investigated further and in more depth.

5.6.3 Outage Duration

Outage Duration, also a secondary parameter in FS-1033, refers to the service restoral function. Outage Duration is defined in the Standard as the average value of elapsed user information transfer time observed during a single outage event. This parameter provides an unbiased estimate of the more familiar descriptor, Mean Time to Repair (MTTR).

While Outage Duration statistics have been gathered for various communications links (wirelines, terrestrial radio, microwave, satellite channels, etc.) and entire systems (the Bell Network, for instance), user requirements have not been addressed. It seems important that the military user needs for all applications and traffic acceptance categories should be established.

5.6.4 Outage Probability

Outage Probability is a secondary parameter in FS-1033. According to Table 18, it serves the message transfer function (see also Section 5.4.11) by representing two performance criteria, accuracy and reliability. Outage Probability is said to be a discrete measure of unavailability. It is defined as the ratio of total message transfer attempts that encounter an outage to the total message transfer attempts. It is the same as the probability that a transaction attempt will be aborted by the system either being in or going into an outage. For quick comparisons, Outage Probability is approximately the same as unavailability, which is commonly defined as MTTR/(MTTR + MTBF).

User requirements for Outage Probability are seldom documented. Service requirements, such as availability of 99% (i.e., approximate Outage Probability of 0.01), have been stated (DCA, 1975) for the general AUTODIN II user class that connects to a single PSN via a single access circuit. Or, the availability of 99.95% (viz., Outage Probability of 0.0005) is to apply to those users who connect to two PSN's with one access circuit each. Kirk and Osterholz (1976) address the allocation of unavailability to various segments of reference circuits, such as the well-defined 600-mile (ca. 1000 km) DCS reference channel. Kelley (1977) emphasizes the importance of user requirements as a tool towards "basic ADP system design objectives". Many application types and their projected traffic percentages are given. However, the actual user availability needs are not quantitatively treated. Finally, Feldman, et al. (1979) advance the viewpoint that 99.9% availability, or 10^{-3} Outage Probability, long term requirements are not needed by conventional users and, therefore, are too demanding on the system. Unfortunately, while they do include various military applications and precedence levels in their study, the highest modem speed considered is 9.6 Kb/s. Thus, while many general conclusions may be valid, it is not at all clear what the availability numbers should be for the interactive users of the 56 Kb/s, Mode VI traffic.

5.7 Summary of Proposed Values for the Primary Parameters

Previous Sections 5.3.1 to 5.5.2 have reviewed various issues involved in number assignment to the 19 primary FS-1033 parameters. Some tentative target values for these primary values have been mentioned and their pros and cons discussed. Whereas in many instances, such as for Access Time in Section 5.3.1, entire ranges and arrays of numbers have been scrutinized, several single numbers do appear more prominent. These singular numbers represent the tentative parameter values summarized in Table 38.

When looking at the numbers proposed in Table 38, one key fact should be kept in mind. That is, the proposed numbers have not been validated for the AUTODIN II digital mode. They do not constitute established service, user or system requirements. Rather, the values should be viewed as examples. The values are quite tentative, although they are generally consistent and do fall in the range of what appears needed on one hand, and attainable on the other.

The validity of the numbers suggested in Table 38 should be verified in a future study. Changes should be made wherever necessary.

6. SELECTION OF ANALOG SERVICE PARAMETERS

In this section we examine the methodology which has been used for evaluating the performance of voice communication channels during the information transfer phase. Unfortunately, we find little standardization of parameters or the techniques used to measure these parameters. The recent advances in microprocessing

Table 38. Tentative Values for the Primary FS-1033 Parameters

Function	Performance Criterion		
	Efficiency or Speed	Accuracy	Reliability
Access	1. Access Time 0.10 s (Mean) 0.15 s (0.9-Perc.)	2. Incorrect Access Probability 10^{-10}	3. Access Denial Probability 10^{-3} (at 0.3s)
Bit Transfer	4. Bit Transfer Time 0.5 s	5. Bit Error Prob. 10^{-10} 6. Bit Misdel. Prob. 10^{-11} 7. Extra Bit Prob. 10^{-11}	8. Bit Loss Probability 10^{-11}
Block Transfer	9. Block Transfer Time 0.5 s	10. Block Error Prob. 10^{-9} 11. Block Misdel. Prob. 10^{-9} 12. Extra Block Prob. 10^{-10}	13. Block Loss Probability $3 \cdot 10^{-11}$
Message Transfer	14. Bit Trans. Rate 8510 b/s 15. Block Trans. Rate 8510/n* blocks/s 16. Bit Rate Eff. 50% 17. Block Rate Eff. 50%		
Disengagement	18. Diseng. Time 0.05 s (Mean) 0.10 s (0.9-Perc.)	19. Diseng. Denial Probability 10^{-3} (at 0.15 s)	

*n = number of bits per block.

offer ways of obtaining new objective measures that satisfy the criteria of reliability, repeatability, useability, and system independence, and that are user oriented in a sense discussed in this section.

Some of the parameters already defined for the digital mode of operation are applicable to the analog mode with an appropriate interpretation. Before selecting the new parameters which are applicable to the transfer phase we present a hypothetical example using access time and access denial to illustrate the procedures for user orientation. In each example, we will assume that a system equivalent to the present direct dial telephone system is available, and that all user groups are familiar with its use. Also, since access time is well defined, we assume it can be measured. Recall that access time is defined only for successful completion of the call.

In the first example, we assume a very sophisticated user group that knows the values of access time it requires. In this case, a user poll is conducted which results in data as illustrated in Figure 50. In this case, the parameter "access time" has a direct relationship to the user response.

In the second example, we assume a less knowledgeable group of listeners, but one which is still restricted. An experiment is established wherein the access time can be varied. The user group is asked to score the acceptability of the system for each of the access times on the basis of 0 to 100%. The resulting information would appear as in Figure 51. In this example, access time is still a directly related parameter, but additional statistical variability is introduced.

In the final example, we assume that the user group is not available for polling, but that they are all using a system on which a number of different measurement capabilities are available. One such parameter is the time for user abandonment after completion of dialing but before any system action (such as a busy signal, a ringing tone, etc.) occurs. These data might appear as in Figure 52.

This parameter would usually be counted in the access denial statistics, but it gives important information about one of the factors contributing to the access time. For example, it indicates that delays between dialing and system response which are greater than T_0 will result in 100% access denial. Thus, in this example we not only have the added statistical variability over example 1, but we also have only indirect evidence about access time.

In each of the above examples, a relationship is established between a measurable system parameter and a user interpretation or evaluation of the effects

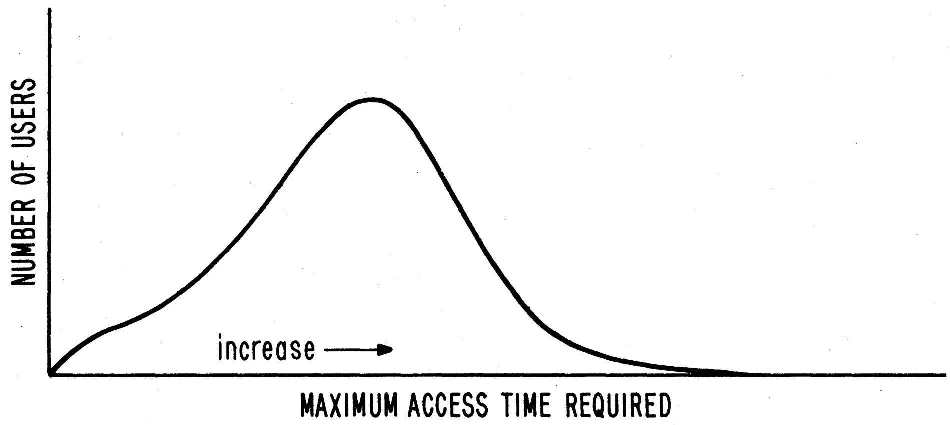


Figure 50. Sample of user estimates of required access time.

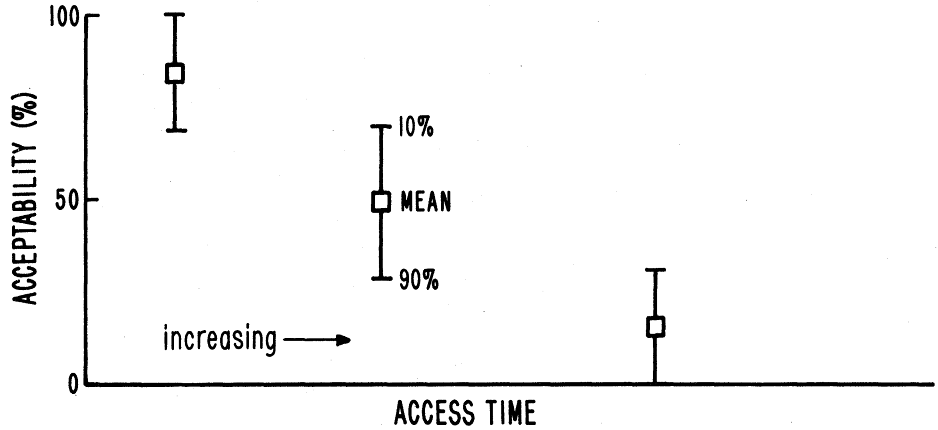


Figure 51. Sample of user scoring of acceptability.

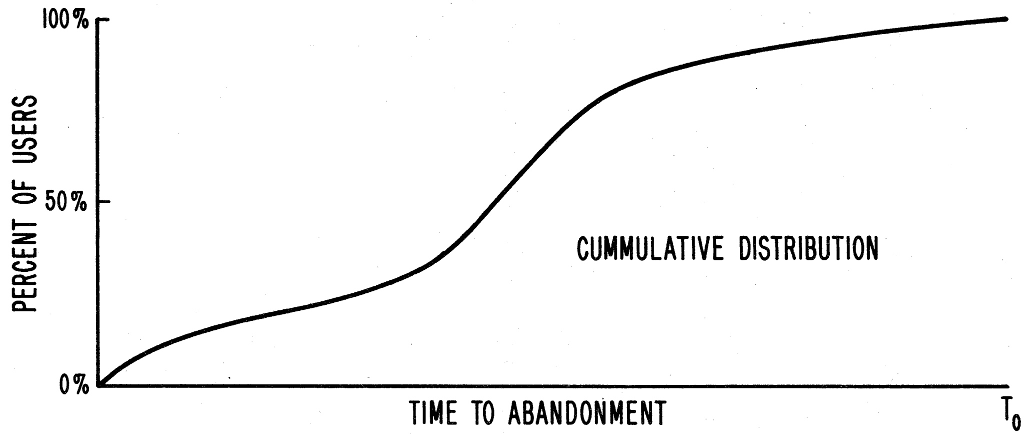


Figure 52. Sample of the cumulative distribution user abandonment times before any system action.

of different parameter values. Thus, the system designer has ranges of values to choose from in an attempt to create a balance between costs, available technology, and numbers of unhappy users.

The relationship between the objective measures and the user applications and interpretation are shown in Figure 53. It is imperative that the objective performance measures correlate well with the human subjective tests and it is desirable that the algorithms for the measures can be implemented simply and quickly. This makes the measure set a good tool. From the military side, (see top of Figure 53), there are two factors discussed earlier in Sections 2 and 3, the various military service classes and their unique user service types and performance requirements. The objective performance measures help to translate these performance needs, with the aid of the correlated subjective test data base, into system planning requirements. This, the fourth arm of Figure 53, shows potential application to the design of military networks. The exact relationship between the arms of the diagram is uncertain at the present time. These uncertainties must be removed in the future by establishing an appropriate data base.

The remainder of this section is designed to select a parameter set for the transfer phase for voice service.

Section 6.1 provides an overview of the problem, Section 6.2 describes certain promising classes of parameters, and in Section 6.3, we select five of these parameters as the candidate set for performance specification and measurement. Section 6.4 outlines procedures for relating the parameters to user opinion.

6.1 An Overview of Voice System Performance Parameters and Measures

The purpose of this section is to define and/or justify criteria for choosing parameters for specifying and measuring the performance of voice communications systems. The methodology will start with a listing of some of the measures and methods of measurement currently used or proposed. When it provides insight, the evolution of the measures will be discussed.

As will become apparent, the paucity of good definitions presents a tremendous obstacle to the understanding and acceptance of voice performance measures. For the purposes of this section it will suffice to emphasize occasionally that a term is undefined, or perhaps overdefined (when several definitions are used). To obtain this emphasis, liberally excerpted portions of other research (with appropriate referencing) are presented.

It is seen that the large number of subjective measures makes it difficult to select a small set. Instead, several classes of subjective measures will be selected

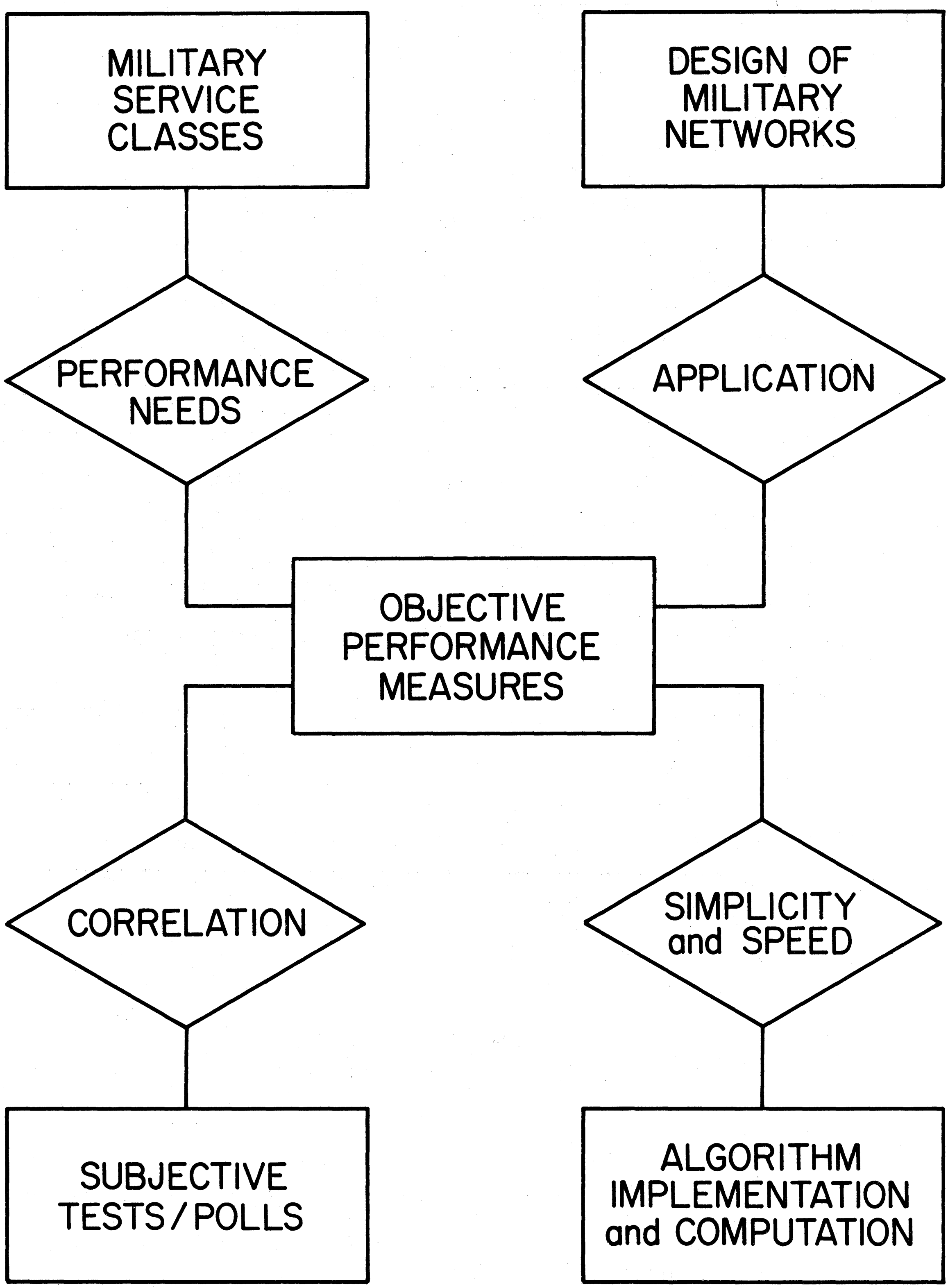


Figure 53. Block diagram of the relationships between objective measures, user applications, and user interpretations.

and each class will be put into correspondance with a single objective measure. Thus, for example, many subjective intelligibility measures will be made to correspond, through calibration, to a single objective measure. This procedure allows the user group to select the subjective measure best reflecting their needs.

6.1.1 Status of Definitions

Webster (1960) defines intelligibility as:

> in•tel•li•gi•bil•i•ty (in-tel'i-jə-bil'ə-ti), n. 1. the quality or fact of being intelligible; capability of being understood; clarity. 2. [pl, INTELLIGIBILITIES] (-tiz) . something intelligible.
>
> in•tel•li•gi•ble (in-tel'i-jə-b'l), adj. [ME.; L. intelligibilis, intellegibilis < intelligere; see INTELLECT], 1. that can be understood; clear; comprehensible. 2. in philosophy, understandable by the intellect only; conceptual.

Typically, definitions from Webster are too vague to be of value for engineering purposes. A definition becomes useful in the engineering sense only after a widely accepted method of assigning numerical values is incorporated into it. In this subsection we look at the reasons why an engineering definition of intelligibility is not available. Attempts have been made to assign numerical values through two types of testing: subjective testing and objective testing.

Subjective Testing

Inherent in the definition of intelligibility are the human users and their judgements. Consequently, most attempts to define intelligibility begin with a method which relies on evaluations of the outputs from one or more listeners. Difficulties arise immediately in the selection of speakers, listeners, and text to be judged and the method of judging. By the time these difficulties are overcome, the testing procedures are complicated, expensive, and (more to the point) out of reach of most of the users who desire test results. Consequently, wide acceptance of any single definition is almost certainly impossible.

Additional barriers to user acceptance are created when selected speakers and listener panels are used to achieve repeatability of results. The user retorts, "How can these results apply to my system which cannot select speakers and listeners?" Much of this reluctance on the part of the user to interpret the measurements for his system would probably disappear if the measurements were cheap and easy to make.

Further confusion is introduced by adding modifiers, e.g., isolated word intelligibility, sentence intelligibility, message intelligibility, etc.

A more basic problem than some of the above is in the form of disagreements over whether to measure "intelligibility" directly or other parameters related to intelligibility, and then calculate "intelligibility."

Finally, if all of the above problems were resolved, the user would say "Intelligibility is not what I'm interested in. I'm interested in quality, naturalness, speaker recognition, user acceptance, and annoyance factors".

Objective Testing

One of the first objective measures used was the signal-to-noise ratio (S/N). Note that S/N is not precisely defined here. For some classes of systems and some measures of S/N, this gives a very good indication of the voice performance, particularly when experience is obtained in interpreting it. S/N is easy to measure, and for this reason it is used as a performance indicator, many times in situations where it is not applicable. Certainly, the articulation index (AI), which is formed from a weighted average of the S/N over a number of frequency bands (most commonly 20), is a very good measure of intelligibility for systems where the signal degradation is entirely caused by additive white Gaussian noise.

Advances in computer technology have revived interest in more complicated objective measures. These approaches can be classified into three broad categories; (1) short term S/N measures, (2) spectral measures, and (3) prediction analyses measures. The relative strength and weaknesses of these measures will be discussed later in the section. Certain of these measures show sufficient promise that they should be the prime candidates for the parameter definition for voice performance.

6.1.2 Other Views

To examine partially the background of voice testing, it is interesting to examine the following quotations from the work of others.

The first six quotations are taken from works at Bell Telephone Laboratories. Note that each gives a unique testing method.

Munson and Karlin (1962)

> "This exploratory paper describes a modification of the paired comparison technique for deriving a one-dimensional scale for rating speech transmission systems on the basis of listener preferences. The numbers of the scale, which run from 0 to 100, are called "Transmission Preference Units" (TPU), and are intended to be used to evaluate any speech transmission system, regardless of the noise or distortions encountered, provided the system is less preferred than the reference condition, namely, real speech at 1 m. If the TPU ratings for two transmission systems are known, it is believed that the difference can be used to predict the percentage of users who would prefer the system with the higher rating.

Subjective listening experiments have been conducted with a small group averaging seven observers for rating a large number of speech transmission conditions on the TPU scale...."

Sen and Carroll (1973)

"For each test condition, 31 subjects heard a 30-s simulated telephone conversation several times, some with low crosstalk and some with high crosstalk level. At the end of each test condition, the subjects rated the transmission quality on a 5-point scale scored, from excellent to unsatisfactory, as 5, 4, 3, 2, and 1, respectively.

"Sixty-seven subjects, all employees at Bell Laboratories, Murray Hill, served as subjects...."

Agrawal and Lin (1974)

"The Measurement of speech intelligibility is of interest to evaluate a speech communication system or the effectiveness of an intelligibility enhancement scheme. It is often reported that a particular system is capable of speech transmission at X bits/s, but it is seldom reported as to how intelligible the system is to the sounds representing normal English conversation. An American National Standards Institute (ANSI) test ... on measuring monosyllabic word intelligibility exists in addition to many other nonstandard tests. Typically, test samples are spoken by a talker, while listeners write them on answer sheets. The percentage of samples correctly recorded is the intelligibility score. Presenting these tests with various lists to many listeners (for statistically stable scores) and evaluating their responses is laborious and somewhat impractical because of time limitations. This probably is one reason why many systems are not subjected to formal intelligibility testing."

Crochiere et al. (1978)

"With the advent of many new speech coding techniques, there has been considerable interest in predicting subjective performance of speech waveform coders based on objective measurements (usually by computer simulation) on the input and output coder waveforms. In this paper a number of techniques and issues involved in the formulation of such measures are examined and an effort to draw together common points of view is made....

"One alternative to using s/n to evaluate coder performance is to resort to extensive subjective tests for comparing coders, but this is generally costly and time consuming. Furthermore this provides little or no insight into how to optimize the design of the coder, or to design new coders. Thus a strong need exists in speech coding for objective performance measures which can do a better job at characterizing subjective performance than the conventional s/n."

Daumer and Cananaugh (1978)

"The subjective tests described in this paper were all conducted as listening-only tests (not conversational). The subjects listened to prerecorded speech and voted on the perceived quality. Details concerning the test facilities, selection of subjects, test circuitry, and test administration are covered in this section...."

"The subjective tests were conducted in an acoustically treated test room containing 11 cubicles permitting up to 11 subjects to be tested simultaneously. Each cubicle contains a handset over which test conditions are heard and a keyboard with five keys labeled "excellent," "good," "fair," "poor," and "unsatisfactory," which is used for registering the vote for each test condition. Associated with the keyboard are red indicator lights which are lit during the presentation of a test condition and green indicator lights which are lit to indicate the period for voting on the test condition.

Subjects were selected from employees in various job classifications and age groups at Bell Laboratories in Homdel, NJ. The sheer number of test conditions dictated that the total test program be divided into five tests. All the tests were administered independently of one another using different subjects...."

Tribolet et al. (1979)

"Digital recordings of sentences spoken by four talkers (two male and two female) were processed by each of the 12 coders. The processed utterances were equalized to the same mean power to eliminate loudness differences. Two analog test tapes were prepared that contained different permutations of four random orderings of the 12 coders. The talkers were assigned in a balanced design so that each coder was represented by the speech of a different talker in each of the random orders. Since each of the four talkers had recorded a unique set of eight sentences, the sentences were randomly assigned and none occurred more than twice.

"Students from the junior and senior classes of local high schools served as paid subjects. They listened to the processed speech binaurally over Pioneer SE 700 earphones while seated in a double-walled sound booth. Sixty-five subjects judged the 48 coded sentences (4 coders x 3 bit rates x 4 talkers). They were asked to rate the quality of each sentence on a scale from 1 to 9, using a 1 to represent the worst quality, 9 to represent the best quality, and the numbers between 1 and 9 for intermediate evaluations. Before the test session began, they judged six representative conditions for practice to familiarize them with the task and the range of quality...."

The next three references give samples of the same type of thing from other authors.

Sergeant et al. (1979)

"The ideal speech reception and discrimination test for evaluating components (talker, channel, listener) of a communication chain must meet two basic requirements: it must be a valid sample of the speech the chain carries or is about to carry, and it must make possible an analysis of errors. Not many speech tests approach the ideal in both these regards. Dozens of sentence intelligibility tests have been constructed, but these are always cumbersome to administer and score, and furthermore, even with key word emphasis within the sentences they do not lend themselves readily to error analysis. Dozens of single-word intelligibility lists have been constructed which are quick and easy to administer and score, and make for very precise error analysis, but represent only poorly the speech material of direct interest.

"A major difficulty with the use of sentence intelligibility tests is that the variance among listeners' responses depends very heavily on the match between each listener's condition (experience, intelligence, etc.), and the vocabulary and content of the message. Single-word closed-set response tests, in which all allowable choices are given on the answer sheet, reduce this variance so far as possible. On the other hand, tests with single word lists do not at all sample the acoustic and prosodic transitions between and within words which are so much a part of colloquial speech...."

Wong and Markel (1978)

"Various linear prediction vocoders have been developed in the last ten years .... Some designs have already been implemented in real time for low bit-rate speech transmission .... There is a continuing effort to improve the designs both in performance and efficiency, such as .... As the quality difference between different designs may not always be obvious, some method of evaluation is often desirable or even necessary. Since the 1920's, numerous evaluation methods have been developed to attain one of two general goals: 1) articulation and intelligibility testing ..., and 2) quality or preference evaluation .... In the evaluation of recently developed systems, such as the LPC vocoders, it is often desirable to have a method that will reveal small differences in the system, and also offer useful correlations between the system parameters and the test scores. For intelligibility testing, the diagnostic rhyme test (DRT) has been developed ... to meet these goals. It uses rhyme words pairs which differ by one of six phonemic attributes patterned after the distinctive features ..., and has been demonstrated to provide reliable and economical measurement of consonant apprehensibility. Several surveys on present-day digital vocoders ... have in fact been conducted using the DRT...."

Hanson (1971)

"Through the years a number of intelligibility tests have been developed, e.g., the Phonetically Balanced (P.B.) Word Test, the Fairbanks Rhyme Test (FRT), the Modified Rhyme Test (MRT), and others. All of these tests, however, have certain shortcomings, e.g., lack of sensitivity, difficult to administer, tedious to score and evalute or inability to perform any diagnostic testing. In an attempt to overcome those shortcomings, Mr. John Preusse of the Speech Technique Team, Information Acquisition Technical Area, CADPL, has developed a new test which is called the Consonant Recognition Test (CRT). It is this test that is used in this series...."

Barnwell and Voiers (1979) reflect the philosophy closest to that chosen in this section.

Barnwell and Voiers (1979)

"This effort deals with a set of techniques which can be used for more effective and efficient operational speech quality testing. In general, these 'objective fidelity measures' are computed from an 'input' or 'unprocessed' speech data set, S, and an 'output' or 'distorted' speech data set, S_0, ... The output speech data set results when the input speech data set is passed through the speech communication system under test. Objective measures may be very simple, such as the traditional signal-to-noise ratio, or they may be very complex. A complex measure might use such diverse measures as a spectral distance or other parameteric distances between the input and output speech data sets; semantic, syntactic, or phonemic information extracted from the input speech data set; or the characteristics of the talker's vocal tract or glottis....

"If an objective fidelity measure existed which was both highly correlated with the results of human preference tests and which was also compactly computable, then its utility would be undeniable. Clearly, it could be used instead of subjective quality measures for testing and optimizing speech coding systems. Such tests could be expected to be less expensive to administer, to give more consistent results, and, in general, not to be subject to the human failings of administrator or subject. Such an objective measure would also be very useful ..."

6.1.3 The Changing Nature of Subjective Tests

In this section, we trace the evolution of one of the more popular subjective tests, the Diagnostic Rhyme Test (DRT), and the trends exhibited by one of DRT's chief proponents and principle investigators, W. D. Voiers.

The early Rhyme tests, (Fairbanks, 1958) were derived from studies on consonant discrimination. The Fairbanks Rhyme test used word lists with differing leading consonants. Numerous changes in the word lists, testing techniques and evaluation followed. Major changes were introduced by House et al. (1965) with the Modified Rhyme Test (MRT), Griffiths (1967) [the diagnostic articulation test (DAT)]. Voiers et al. (1965), Voiers (1967), and Voiers (1971) include a continuing series of changes in the DRT.

In the implementation of the DRT, various rating forms were tested to determine a set of "Perceived Acoustic Traits (PATS). Table 39 shows two such rating forms.

More recently, Voiers has extended diagnostic testing in an attempt to determine quality and acceptability of a system. He has developed successively the Paired Acceptability Rating Method (PARM) (Voiers, 1976), the Quality Acceptance Rating Test (QUART) (Voiers, 1976), and the Diagnostic Acceptability Measure DAM (Voiers, 1977). Table 40 shows the rating forms for determining the PATS used with this version of DAM.

The continuous change in subjective testing demonstrates the lack of stability of the testing methods. Unfortunately, no clear-cut goals are established for research on subjective testing. This accentuates our previous contention that no <u>widely accepted</u>, inexpensive and easily administered and interpreted subjective tests are available now. It also appears that no such tests will appear in the near future.

6.1.4 Some Often Mentioned Qualities of Voice Channels.

1. intelligibility
2. acceptability
3. speaker recognizability
4. naturalness
5. annoyance factors
6. quality
7. fidelity

As shown in Section 6.1.3 there is no widely acceptable definition of intelligibility. In general, the other qualities listed above are more remote from an acceptable definition than intelligibility. The one possible exception is speaker recognizability. This is so because of recent advances in automated speaker recognizability algorithms designed for computer implementation. However, there is a large gap between the present thrust of the research and the use of these methods for voice service systems evaluation.

Table 39. Perceived Acoustic Traits Rating Forms

a

Steady	Fluttering	Excited	Calm
Stable	Unstable	Agitated	Serene
Colorless	Colorful	Gliding	Scraping
Monotonous	Dynamic	Smooth	Rough
Foreign	Native	Fast	Slow
Rare	Common	Busy	Resting
Rumbling	Whining	Beautiful	Ugly
Low	High	Clean	Dirty
Unpleasant	Pleasant	Feminine	Masculine
Annoying	Pleasing	Light	Heavy
Gradual	Abrupt	Familiar	Strange
Rounded	Jagged	Usual	Unusual
Loud	Soft	Clear	Hazy
Intense	Mild	Definite	Uncertain
Passive	Active	Uneven	Even
Dragging	Brisk	Irregular	Regular

b

Thumping	Resonant
Breathy	Hissing
Twangy	Tinny
Solid	Closed
Clicking	Smacking
Squeeking	Chipping
Babbling	Gurgling
Snapping	Crackling
Thudding	Dull
Abrupt	Clipped
Throaty	Rich
Hooting	Bleating
Rushing	Gushing
Buzzing	Droning
Hollow	Open
Tight	Tense

Table 40. Rating Form for Determining Perceived Acoustic Traits Used in DAM

<u>The Speech Signal</u>

Fluttering
Twittering - Pulsating

Muffled
Smothered - Low

Distant
Small - Compact

Rasping
Scraping - Grating

Thin
Tinny - High

Unnatural
Mechanical - Lifeless

Babbling
Chortling - Slobbering

Irregular
Spasmodic - Fitful

Nasal
Whining - Droning

Interrupted
Intermittent - Chopped

<u>The Background</u>

Hissing
Simmering - Fizzing

Chirping
Cheeping - Clicking

Roaring
Rushing - Gushing

Crackling
Scratching - Staticy

Buzzing
Humming - Whirring

Rumbling
Thumping - Thudding

Bubbling
Gurgling - Percolating

<u>The Total Effect</u>

Intelligible
Understandable - Meaningful

Pleasant
Rich - Mellow

Acceptable

For certain systems, a "quality" may be defined, for example, toll quality for analog systems having the characteristics of the telephone system. However, this definition is the result of extensive experience with the system, and it does not apply to other systems, even though some other systems are said to offer voice services of toll quality.

The concept of "annoyance factors" probably is composed of a number of separate qualities and therefore it may not be appropriate to list it as a single measure or quality item. It would include, but not be limited to, such things as background type and level, time availability and access time, interruptions, ease of use (e.g., push to talk?), delay, echo, etc. A major effort would be needed to arrive at a useful definition of annoyance. The probability of the success of such an effort appears to be small.

The acceptability of a system is a complicated, unknown function of all of the other qualities. Some attempts have been made by Voiers (1976) to create repeatable subjective tests for a class of systems and a limited class of users to measure acceptability. Since these tests are still evolving, it is difficult to judge their validity and efficiency. In any case, they are still <u>subjective</u> and suffer from the same problems as other subjective tests - they are not widely acceptable.

The qualities naturalness and fidelity are, at this time, so vaguely defined as to be useless in user oriented service assessment.

6.2 Classes of Objective Measures for Determining Voice System Performance

Any system performance measure should satisfy a certain minimal set of criteria. These may be specified as follows. (1) Reliability: this specifies that anyone making the same measurement at different times should get the same answer. (2) Repeatability: this is the property that a measurement be specified sufficiently so that measurements made by one group can be repeated or verified by another group. (3) Useability: this property involves several factors such as time, cost, and complexity. The criteria which we strive for are short test time requirements (e.g., 1 hour for a single measurement, including set up, calibration etc., with shorter time requirements for additional measurements), low recurring costs, and simplicity (e.g., only one instrument needed for all system inputs and one needed for all output measurements). The instrument may be internally complex, but the use of it should be simple. (4) System independence: the parameters chosen should be as widely applicable as possible, and therefore not dependent on the type of system being tested. However, the interpretation of the parameter values may (and probably will) depend on a priori knowledge of the system. (5) User oriented: we list this last, because it depends greatly on the users' experience with their system, the expected service, and with the measurements chosen, but the performance measure must provide significant correlation with subjective measures including various forms of intelligibility, speaker recognition and other voice transmission quality features previously discussed.

Because the voice system affects the performance between the speaker's mouth and the listener's ear, we will design the parameters around a system which has an input (representing the speaker's mouth) and an output (representing the listener's ear) and use these parameters to deduce the essential perceived voice characteristics. The actual measured parameters will consist of various differences between the input and output signals.

In order to make the following recommendations precise we assume that all processing is done digitally. It will become clear that all of the techniques discussed can be implemented using microprocessors. This method of implementation allows changes to be made through software which results in a flexibility not achieved in hardware.

The input signals will be generated from stored, digitized information and synthesized into an appropriate analog form. This permits accurate synchronization of the output signal with a stored replica of the input signal. Consequently, in what follows we assume that the input signal is known to within 1 sample for any sampling frequency discussed.

The output signals will be digitized and processed according to the parameters being measured.

A block diagram of the testing system is shown in Figure 54. The interfaces depend on the points at which the system is being tested. For a telephone input the interface would be an acoustic coupler.

6.2.1 Mathematical Modeling

One of the requirements for the objective measures is that it be simple enough to permit implementation. Fortunately, a mathematical technique known as Linear Predictive Coding (LPC) provides a basis for easily implementing algorithms which result in several classes of measures with potentially effective interpretations in terms of voice performance parameters.

A review of LPC is given in Appendix C. There, we show that, after the initial calculation of a small set of parameters, a number of measures related to

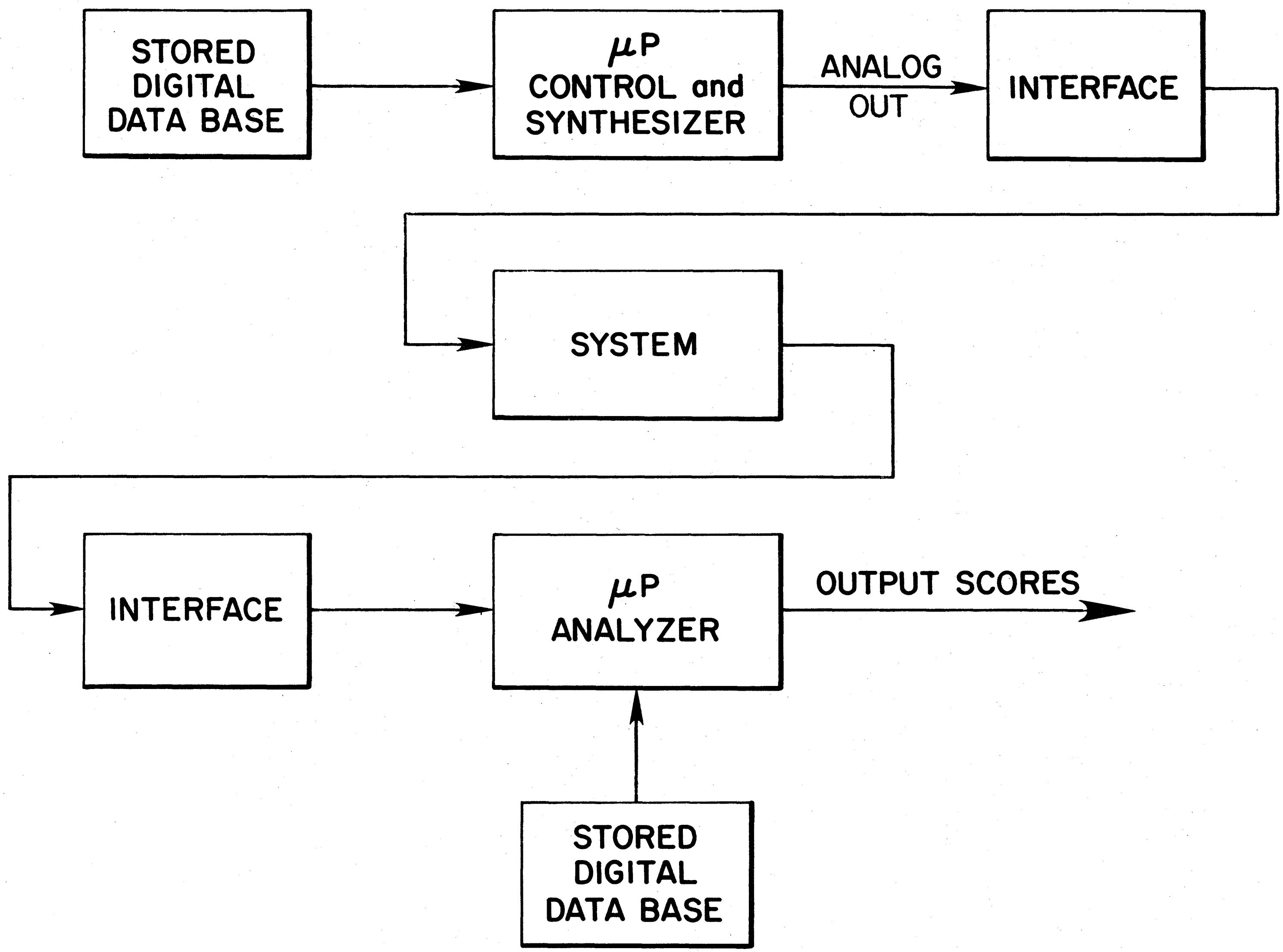


Figure 54. Block diagram of system for performing objective measurements of voice performance.

voice performance are easily calculated. The measures derived in Appendix C relate to short term spectral distances and vocal tract dimensions and have interpretations in the time domain and frequency domain.

Since LPC analysis and synthesis have been implemented in real-time vocoders using microprocessor based techniques, it is clear that the LPC parameter set can be derived using the same techniques. Appendix C describes the simple calculations using this set to derive the measures of interest.

6.2.2 Measures

Some of the measures which we propose are based on the smoothed spectral estimate of equation (C-1), the power spectral ratios (energy ratios) of equation (C-51) and the area coefficients of equation (C-54). The sampling rate, quantizing levels, number of points in a frame, and the number of coefficients computed must be determined for each of the measures. However, we leave these unspecified at this time and only describe the form for the suggested measures.

Spectral measures.

Let

$$SP(n,p) = \left[\frac{1}{N} \sum_{j=1}^{N-1} \left| \log_{10} \frac{H_n(z_j)}{H_n'(z_j)} \right|^p \right]^{\frac{1}{p}} \tag{1}$$

where $H_n(s)$ is the spectral estimate for the nth frame of the input and $H_n'(z)$ for the output and $Z_j = e^{-i\pi\frac{j}{N}}$. The spectral measure proposed is then given by

$$SP(p) = \frac{1}{k} \sum_{n=1}^{k} SP(n,p) \text{ for } p = 2, 4. \tag{2}$$

Barnwell and Voiers (1979) have compared these measures with subjective scores and show some correlation, although they do not appear to be outstanding predictors.

Area coefficient measures.

A second class of measures studied by Barnwell and Voiers (1979) use the form

$$AR(n,p) = \left[\sum_{i=1}^{P} \left[\log_{10} \left| \frac{A'_i}{A_i} \right| \right]^p \right]^{\frac{1}{p}} \tag{3}$$

and the average over the total number of frames

$$AR(p) = \frac{1}{k} \sum_{n=i}^{N} AR(n,p), \; p = 1, 2 \tag{4}$$

These measures exhibit a better correlation with the subjective scores than the spectral measures.

Energy measures.

The energy ratio (C-51) defined for each frame, $(D/E)_n$, has been used in the form

$$ER = \frac{1}{n} \Sigma \log_{10} (D/E)_n \tag{5}$$

by several authors (Sambur and Jayant, 1976; McDermott et al., 1978) (see also Gray and Markel (1976) for related measures). Gamauf and Hartman (1977) used a modification of this which involves testing whether $\log_{10} (D/E)_n$ is between certain confidence limits. They derive a number ElN for each frame where $0 \leq \text{ElN} \leq 1$. Then, using a weighting for each word, an average measure $\overline{\text{ElN}}$ is derived for 50-word phonetically balanced word groups. This normalized energy measure correlates well with the articulation score obtained using trained listener panels.

Signal-to-noise ratios

One of the earliest objective measures used was the signal-to-noise ratio. For analog systems with white Gaussian noise, the articulation index which is a frequency weighted signal to noise ratio is a good predictor of intelligibility, and possibly other subjective parameters [Kryter (1962), Kryter and Ball (1964), ANSI (1969), Hubbard and Hartman (1974)].

More recently, Barnwell and Voiers (1979) have found that a short term signal-to-noise ratio is an excellent predictor of subjective parameters for certain types of system distortion.

When the signal is known, the signal-to-noise ratio can be calculated two ways; filtering the signal to obtain $\frac{S+N}{N}$, or subtracting the signal, $s_n' - s_n$ to obtain

$$\log_{10} \frac{\Sigma (s_n')^2}{\Sigma (s_n' - s_n)^2} \tag{6}$$

Note that in either case the noise may consist of interference and distortion in addition to or in place of the noise. Equation (6) may be difficult to implement in some cases due to gain uncertainties etc. Consequently, in what follows, we assume the quantity $\frac{S+N}{N}$ is obtained by filtering the signal. The signal-to-noise ratio SNR (in dB) may be calculated over a time interval (N points) or over different frequency bands f(j), ($f_j \leq f(j) \leq f_{j+1}$). The notation SNR (N,f) will be used here, where f denotes the band of interest and N indicates the number of samples and $N(i) \leq N_i \leq N(i+1)$. Thus, two classes of measures are defined as

$$SNR(T) = \frac{1}{k} \sum_{i=1}^{k} SNR(N_i, f), \tag{7}$$

where f is the entire frequency band, and

$$SNR(\omega) = \frac{1}{k} \sum_{i=1}^{k} SNR(N, f_i), \tag{8}$$

where N usually represents at least several seconds of data in (8). Barnwell and Voiers (1979) report excellent correlation of SNR(T) with subjective scores for some types of system degradation, and Steeneken and Houtgast (1979) report excellent correlation of SNR(ω) with subjective scores for certain types of system degradation.

6.2.3 Speaker Recognition

Numerous computer oriented methods designed to recognize or identify a speaker have been developed over the past several years. General features of all of these methods include: (a) some stored sample parameters from the speaker, sometimes called the training set, (b) a fixed number of speakers from which a selection is to be made for recognition purposes, (c) a sample utterance from some speaker for which parameters are to be computed for comparison with the training set. The parameter sets are generally derived using either LPC analyses, discrete Fourier Transform (DFT) analyses or cepstral (Luck, 1969) analyses. The computational complexity increases from LPC to DFT to cepstral analysis. The parameter sets used most often compare either the frequency/time structure of the utterances, or parameters such as the area coefficients which have interpretations in terms of the vocal tract dimensions. (We do not consider the method of voice prints directly here, since this is included in the general descriptions given above.) Frequently used parameters, usually used in combinations include formats, pitch, and intensity.

Since our objective is user oriented performance measurement and not speaker recognition per se, we propose the following scheme.

Since our test signals are synthetic, one can generate a training sequence using a single word or a few words, with controlled speaker differences. For example, with (synthesized) speakers a_1, a_2, ..., a_n we can design the analyses and speaker differences so that only a small difference can be measured between adjacent speakers a_i , a_{i+1}, but a large difference exists between a_1 and a_n under clear-channel conditions. Under clear channel conditions, using the same a_1, ..., a_n set for comparison, 100% correct identification (or recognition) would be achieved. However, under other channel conditions less than 100% correct scoring would be achieved.

Since, with few exceptions, speaker recognition algorithms have not been used or tested for different channel conditions, the training set and the specific algorithm to be used remains to be determined.

6.3 Specific Measures

In the previous section we have outlined several classes of objective parameters which have been shown to have utility as predictors of various types of subjective voice performance measures, and some other parameters which, with careful design should predict other voice performance characteristics. The principal user oriented performance characteristics which are considered are (1) intelligibility, (2) speaker recognition, and (3) user acceptability.

For intelligibility testing we recommend four objective measures:

(1) The normalized energy measure developed by Gamauf and Hartman (1977), because it shows good correlation with subjective scores over a wide range of system conditions. This measure is a modification of the energy ratio (5) which involves testing whether log $(D/E)_n$ is between certain confidence limits to derive a number between zero and one. This is then averaged over all frames.

(2) A short term signal-to-noise ratio (Barnwell and Voiers, 1979), because it shows excellent correlation with subjective scores over certain limited ranges of system conditions. This is a variation of the measure given in (7).

(3) A band weighted signal-to-noise ratio [e.g., Steeneken and Houtgast (1979)], because it shows good correlation with subjective scores over the range of system conditions for which it has been tested. This measure uses a modification of (8) with a specified set of input signals.

(4) The log-area ratios (Barnwell and Voiers, 1979) because it is easily computed from parameters derived when calculating the normalized

density and has shown fair correlation with subjective scores. This measure has been tested using the form given in (4).

For voice recognition, we have found few subjective measurements for comparison with objective scores; furthermore, no objective measures have been developed. Numerous techniques are available for constructing such a measure [see for example, Rosenberg (1973), for a general review, and Atal (1974), Luck (1969), and Atal (1976), for a description of some of the techniques]. Rosenberg (1973), using formant, pitch, and intensity as the comparison basis for the computer algorithm, compared computer speaker verification with human speaker verification. Although the results are not directly related to the present emphasis, the fact that the computer scored 98% correct and the humans 96% correct indicates a favorable potential for a computer technique. Further research is necessary to determine the appropriate techniques for this application. However, some of these methods use LPC analysis, and can be easily implemented using information obtained when computing the measures for intelligibility, suggesting that these methods be explored first.

For user acceptance, the measures given for intelligibility have also been shown to be correlated with subjective parameters which have been shown to be correlated with user acceptance [Barnwell and Voiers (1979) and Voiers, (1976)] over a very limited set of conditions.

The three parameters, intelligibility, voice recognition, and acceptability, chosen as voice performance descriptors, are not independent. Thus, a system with low intelligibility usually would not have good voice recognition properties and would not be acceptable. On the other hand, a system may be unacceptable because of long delays in the access phase or the transfer phase even though the intelligibility and voice recognition are outstanding. The objective measures are chosen to quantify these parameters. Table 41 shows the relationship of objective measures to the subjective measures. The primary application means that the objective measure alone is a good predictor of the subjective parameter, while the secondary application implies that other measures are also needed to predict accurately the subjective interpretation.

Table 42 gives an indication of the ranges of applicability of the various measures that have been tested. The wide range of systems indicates that the measures have been compared with subjective data over analog, wide bandwidth digital systems, narrowband digital systems, and a range of interference or jamming environments. The wide range of degradation indicates that, for example, the intelligibility scores have been over the range from 0 to 100%. The restricted range

Table 41. Principal Application of Objective Measures to Subjective Interpretation

Objective Measure \ Subjective Interpretation	Intelligibility	Acceptance	Speaker Recognition
Normalized Energy	1	2	*
Log Area Ratios	1	2	*
Short Term S/N	1	2	*
Band Weighted S/N	1	2	*
Speaker Recognition	---	---	1

(1) Primary Application
(2) Secondary Application
* Applicability not known.

Table 42. Summary of Comparisons of Objective Measures with Subjective Measures

Objective Measure \ Subjective Interpretation	Intelligibility		Acceptance		Speaker Recognition	
	(1)	(2)	(1)	(2)	(1)	(2)
Normalized Energy	W	W	N	N	---	---
Log Area Ratio	W	W	W	W	---	---
Short Term S/N	R	S	R	S	---	---
Band Weighted S/N	R	W	R	S	---	---
Speaker Recognition	---	---	---	---	N	N

(1) Range of Systems

(2) Range of Degradations

Legend: W = Wide; R = Restricted; S = Small; N = Negligible
--- = None.

indicates that the measure has been compared with subjective ratings only for classes of systems where it is applicable. Thus, we see that additional data are needed to complete the comparisons.

6.4 Relation of Objective Measures to User Opinion

Very little data are reported in the literature which give a numerical comparison between objective and user evaluation of a system. This is primarily because user opinion is obtained from very informal polls (or tests). In addition, unless restrictions are placed on the user group selected for the poll, the spread of data is large, the reliability uncertain, and the repeatability nil. However, valuable information can be obtained from such polls. A common type of such a poll simply asks which of two systems is preferred. A majority preference for one system over the other can provide the basis for system selection or modification.

In what follows, we report on one set of data [see O'Brien and Busch (1969), and Gierhart et al., (1970)] for which comparisons are available between objective scores and listener polls. This example is not a good example for several reasons enumerated below. It is cited here because it is one of the few examples available.

6.4.1 The Objective Measures

The two objective measures reported are the long term signal-to-noise ratio (S/N) measured at the intermediate frequency (IF) and SCIM, (Kryter and Ball, 1964), which is a band weighted signal-to-noise ratio measure, measured end-to-end, which approximates the articulation index (AI). For the systems and noise conditions reported, SCIM is a good objective measure, but is not one of the recommended measures because of its sensitivity to system changes.

6.4.2 The Opinion Poll

The biggest problem with this example is that the listener poll is too structured. For this poll, the speakers and listeners were chosen from a select group of air traffic controllers, the tapes were recorded in a noise-free, soundproof room, and the listening tests were done in the soundproof room. Thus, the results would not apply to the more general situation in which the messages would originate with a pilot (perhaps with an accent), in a noisy airplane cockpit, and would be received by an air traffic controller in a less than quiet center. Sample messages are shown in Table 43. The entire message list consisted of 216 messages. The scoring was done by having the listener circle the appropriate response for the set in Table 44.

Table 43. Sample Arrangement of ATC Messages

Item #	Message
1	Cleared to 170, AA 306 leaving 370.
2	There's some cells up here. Anybody been detouring anywhere?
3	Roger right heading 340° and 2244, 294.
4	Eight twenty-two is cleared to two six zero to Victor one three zero, we're leaving.
5	914 X-Ray, descend and maintain 6.0 thousand, we're leaving 10.0.
6	Ok we'll hold 330 we can expedite our descent when we get by these clouds.
7	Cagey 24 heading 218 degrees making good a track of 212 for CYN correction for ORF.

Table 44. Sample of ATC Message Test Score Sheet

a. Message is completely understood.

b. Able to understand most of the message.

c. About half (1/2) of the message could be understood.

d. Could only understand a small amount of the message--would request a repeat.

e. Completely unable to understand the message, would definitely request a repeat.

1. a b c d e
2. a b c d e
3. a b c d e
4. a b c d e
5. a b c d e

.
.
.

6.4.3 The Data

Keeping in mind the shortcomings, we present the test data as a sample application of the methods.

Figure 55 shows the relationship between Articulation Index (a band weighted signal-to-noise ratio objective measure) and several subjective intelligibility scores. Note that the sentence intelligibility scores are not as sensitive a measure as the other tests. This is typical of tests which have redundancy in the test material. What is not shown here, but is in some of the figures to follow, is the spread of data around the curves. Almost always, the data spread is much larger around the less sensitive measures.

To return to our data sample, Figure 56 shows the relationship between S/N (for AM) and the SCIM score (AI) and an Air Traffic Control (ATC) message evaluation. The data spread would increase from negligible for S/N, to slight for SCIM, to large for the ATC messages as shown by the bars which represent ± 1 standard deviation.

Figure 57 shows the same information for FM systems. Figures 58 and 59 present the subjective scores plotted vs. the objective (SCIM) scores for the AM and FM respectively.

6.4.4 Sample Use of the Data

To clarify the concepts in this section we use the following hypothetical scenario:

An AM system is presently being used by trained air traffic controllers in an ideal environment (no background noise, no fading, and only Gaussian noise as a system limitation). An engineer is assigned the task of specifying an FM system which will just give 100% intelligibility for ATC messages.

The engineer would first consult the data in Figure 55. Although this is not directly applicable, the similaritiesbetween sentence intelligibility and ATC message intelligibility are sufficient to allow use of the results in Figure 55 as a first approximation, keeping in mind that the ATC messages are slightly more redundant. Figure 55 gives almost 100% intelligibility for AI = .4 (SCIM = 40%).

6.4.5 Summary

In this section we have examined a number of user oriented parameters for the transfer phase. For these parameters to be useful, it is necessary that there exist a corresponding set of system oriented parameters and methods for relating or calibrating the two sets. For voice systems, this required identifying and

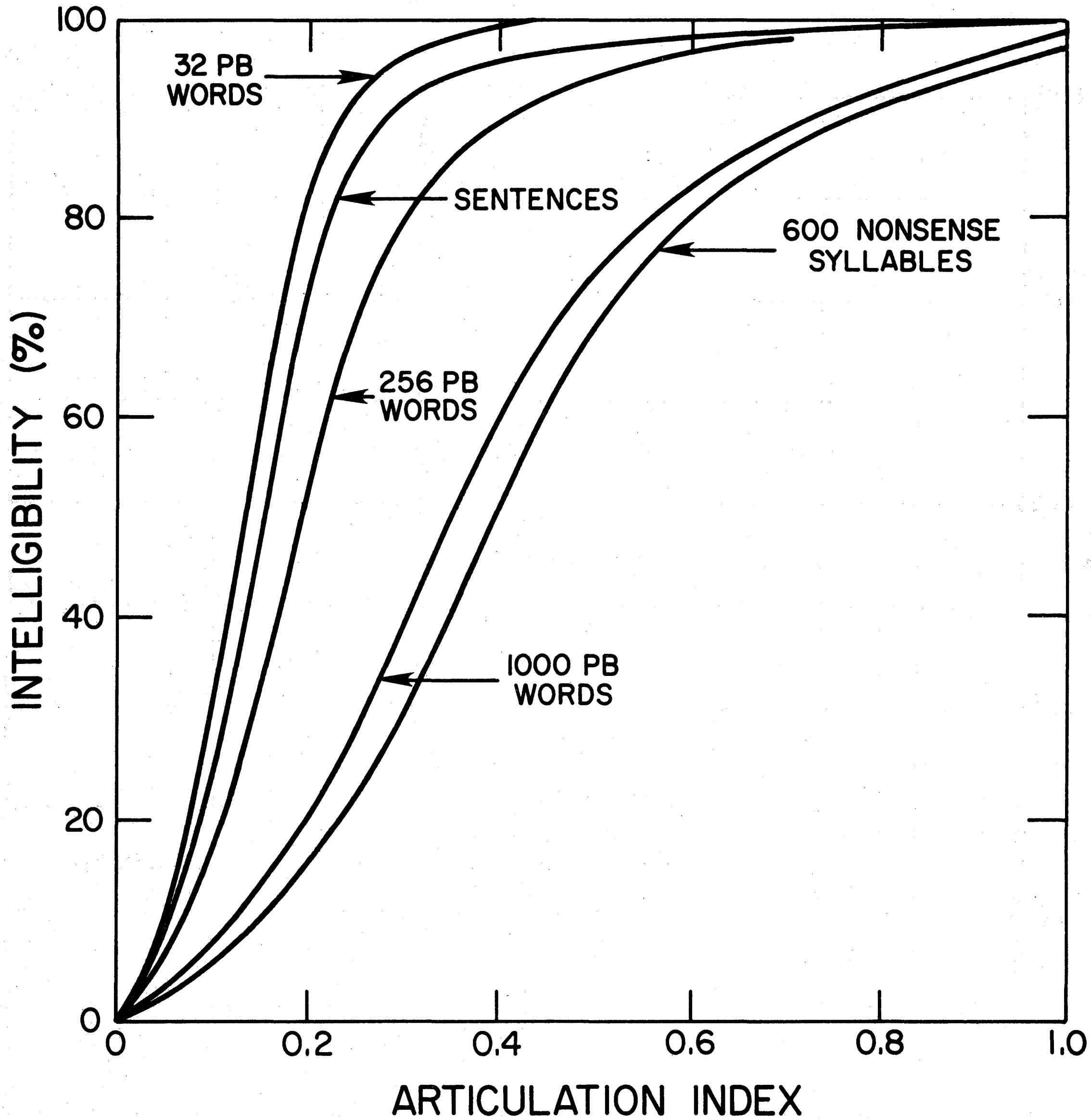


Figure 55. Comparison of different subjective intelligibility measures.

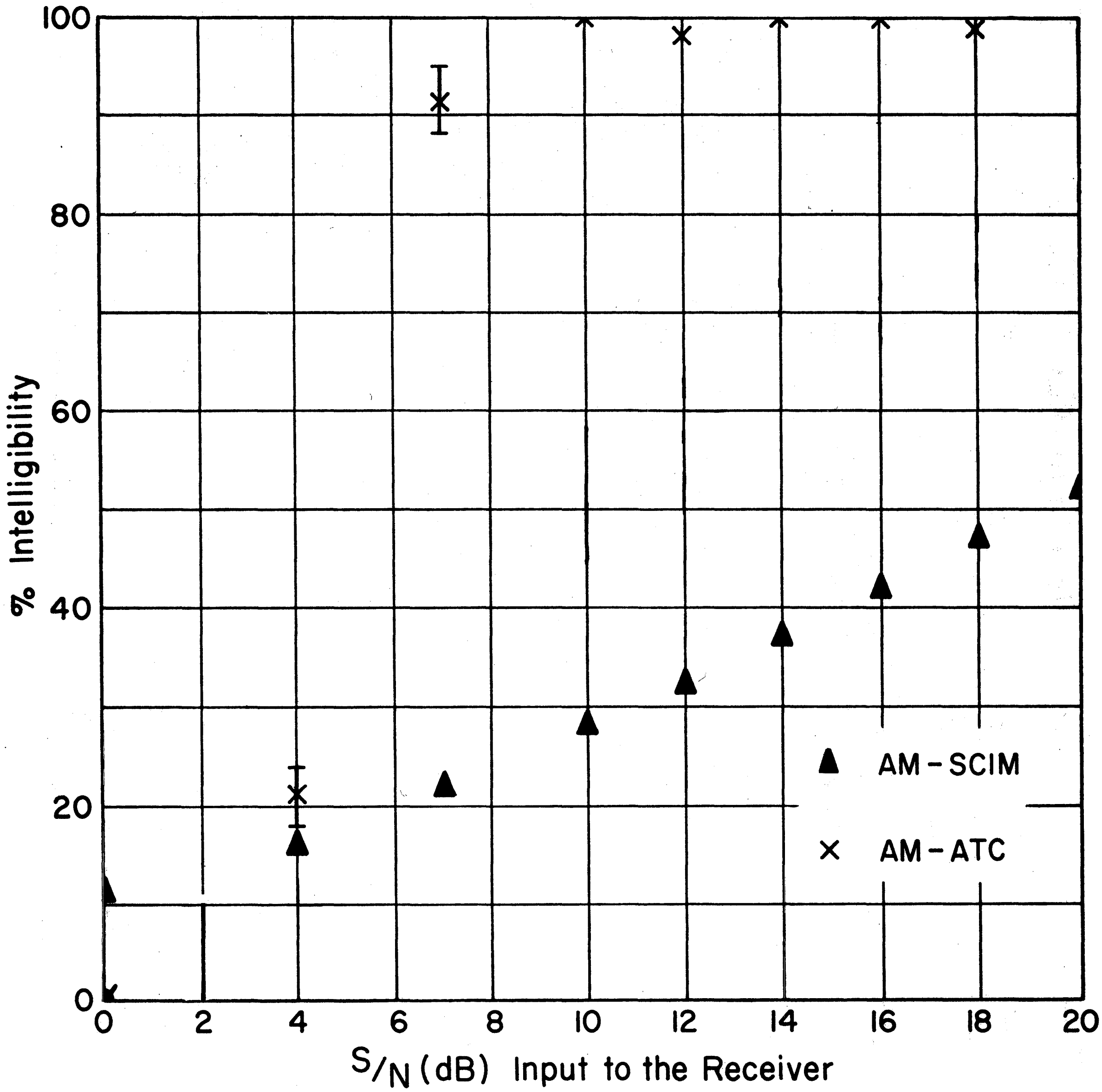


Figure 56. Performance of an AM system in the presence of Gaussian noise.

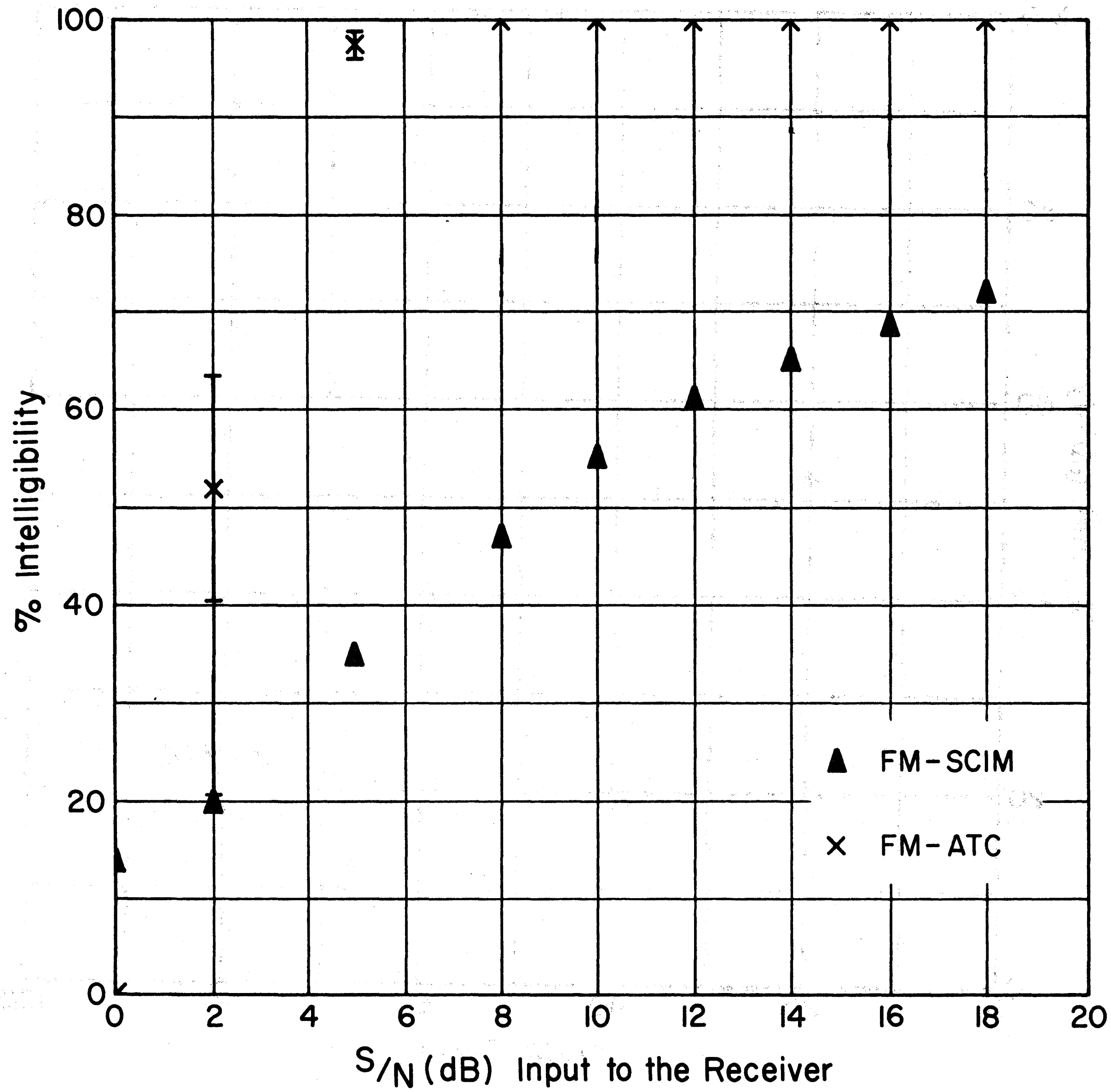


Figure 57. Performance of an FM system in the presence Gaussian noise.

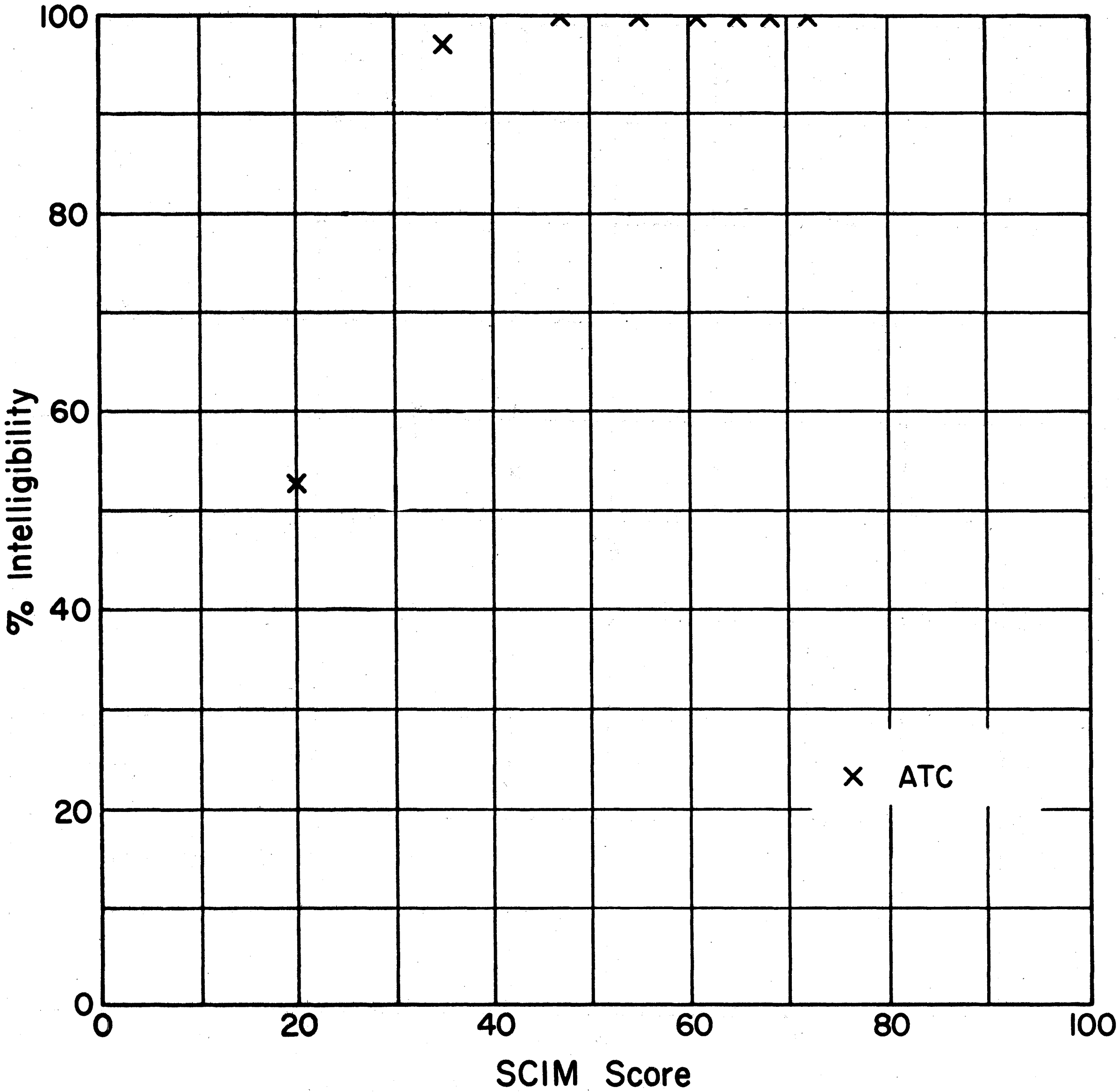


Figure 58. ATC message scores compared to SCIM for an AM system.

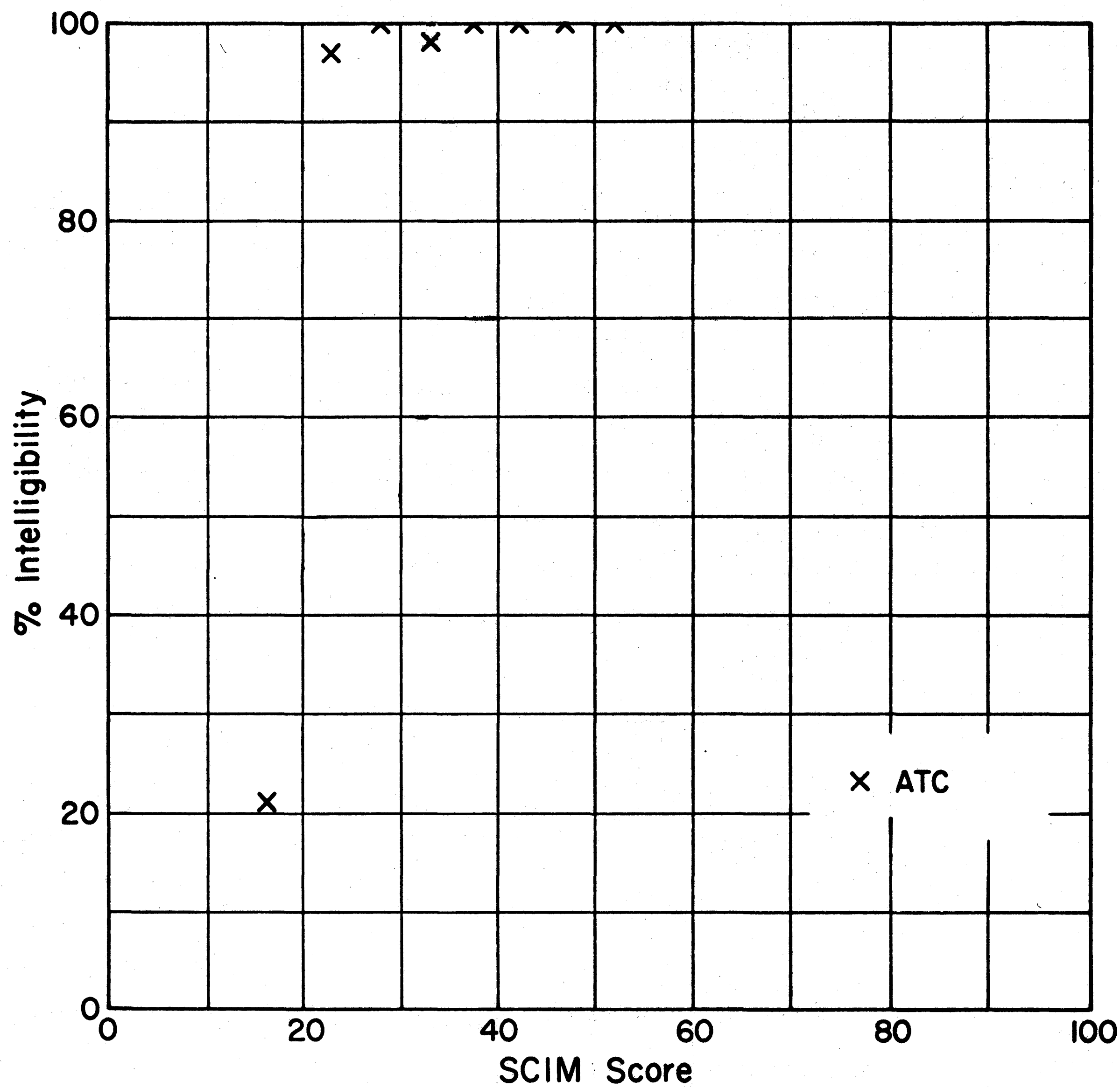


Figure 59. ATC message scores compared to SCIM for an FM system.

defining these system oriented parameters. The selected user oriented parameters are those which can be calibrated in terms of the system parameters.

Table 45 lists the parameters for all phases, including those defined in previous sections which are pertinent to voice systems. Note that some of the user parameters also describe the system parameters.

The three parameters intelligibility, acceptability, and speaker recognition are not completely specified in order to provide flexibility. Thus, if isolated word intelligibility is important to one user group and sentence intelligibility is important to another group, no conflict is present. Both can be related to corresponding objective measures, which are completely specified.

7. ASSIGNMENT OF VALUES FOR ANALOG SERVICE MODE

As described in Section 6, one of the important steps in determining the user oriented parameter values is using the information available from previous testing of either identified systems or other user groups. Such values are reported in this section.

In Section 6 we selected the user oriented parameters for voice service, with the particular application to the 16 Kb/s CVSD system as a guide, but with the goal of wider applicability while still keeping the parameter set to a minimum. Here we outline the general procedure for obtaining values.

Figure 60 shows a block diagram of the procedures.

The first step is to limit the problem by identifying the constraints imposed on the system, such as bandwidth limitations, types of propagation, whether digital or analog transmission is required, etc., describing as precisely as possible the class of systems and operating conditions to reduce the amount of measurement needed. For example, if all transmission must go through a satellite link, round trip delays of less than 500 ms need not be considered in designing the listener poll. Specifying a 16 Kb/s CVSD or one using a specific algorithm further limits the ranges needed for the measurements.

The second step is defining the user class, or classes. This should include an estimate of the amount of training or experience the users will have prior to their normal pattern of system use. This information should be used when possible for designing the user poll, and always in interpreting the results of the user poll.

The results of previous listener polls (with similar user groups, if available), or previous system measurements, should be used to restrict further the measurement range when possible.

Table 45. Analog Service Performance Parameters

Part A - Primary Parameters
1. Access Time
2. Incorrect Access Probability
3. Access Denial Probability
4. Intelligibility
(a) Normalized Energy
(b) Log Area Ratios
(c) Short-Term S/N
(d) Band-weighted S/N
5. Acceptability
(a) Normalized Energy
(b) Log Area Ratios
(c) Short-Term S/N
(d) Band-weighted S/N
6. Speaker Recognition
(e) Computer Speaker Recognition
7. Delay
(f) Round Trip Delay
8. Disengagement Time
9. Disengagement Denial Probability
Part B - Secondary Parameters
10. Service Time Between Outages
11. Outage Duration
12. Outage Probability
Part C - Ancillary Parameters
13. User Access Time Fraction
14. User Delay Time Fraction
15. User Disengagement Time Fraction

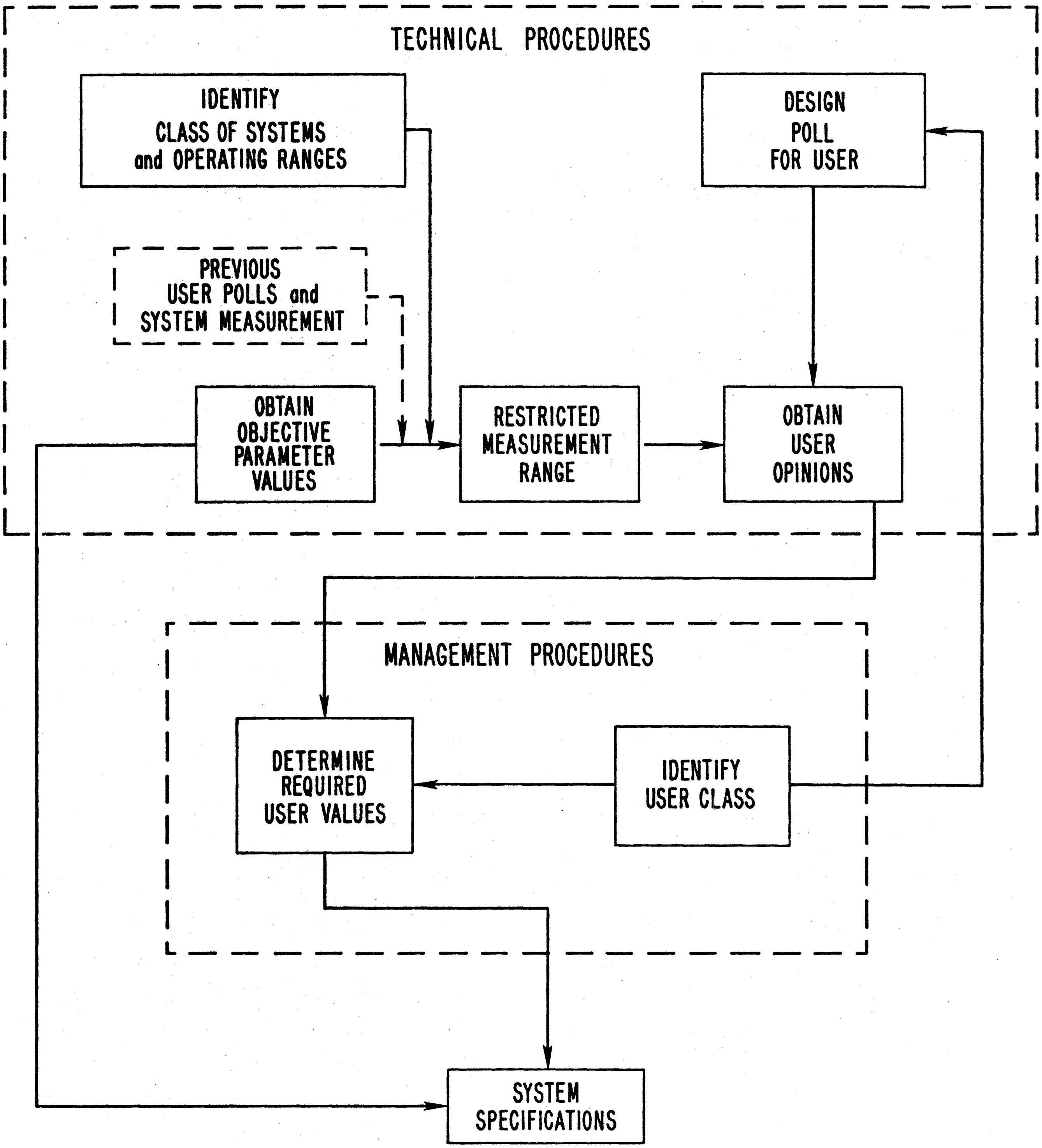


Figure 60. Diagram of the procedure for obtaining user oriented parameter values for system design or testing.

Finally, the objective measurements and opinion polls are conducted under identical conditions to establish the proper calibration.

The identified system is the 16 Kb/s CVSD voice system operating through a Harris modem (Perkins and McRae, 1978) over the AUTOVON network. Consideration will be given to the operation of the voice system in either a clear or encrypted mode.

In previous sections, the definitions for the different parameters implicitly assume a degree of statistical independence. For example, the probability of access denial for one event is assumed independent of the probability for another event. Moreover, it is assumed that the values of the parameters for one phase are statistically independent from those of other phases. When these parameters become dependent, the problem becomes one of system capacity and/or traffic flow. The values reported here do not distinguish dependent and independent events and should be applied accordingly.

For voice circuits the parameter values are usually determined by varying a single parameter and determining its properties. The user perceives the effect of all of the parameters. Thus, although measurements have been made over ranges of voice quality with delay held constant, and over ranges of delays with the quality held constant, no measurements have been made varying both simultaneously.

Because most of the users of the military systems will also be users of the commercial telephone system, their opinions will be influenced by this experience. Thus, the values which are provided here for the commercial direct distance dialing (DDD) service provide an anchor value for specifying or determining the relative performance of other systems. Military users, even though specialized, will subconsciously use this anchor point in specifying their needs for voice communications. Therefore, the values for the commercial system, although not directly applicable, are pertinent to the selection of ranges of user parameter values for any military voice network. One very great difference in the specifications of the commercial service and the military service, is the amount of time a level of performance is achieved.

Since the values given for the subjective parameters are in terms of percentages or in terms of scores which range from 0 to 100%, it is very convenient to normalize the objective parameters to the same range. For the five objective parameters specified in Section 6 the normalized energy ratio and the band weighted S/N have been normalized to the range from 0 to 100%. It should be easy to normalize the short term S/N and the speaker recognition measures to this same range; however, the log area ratio will need further work in order to obtain a normalized range.

Each of these percentages may take a different meaning for the different subjective measures or for different user polls. For example, intelligibility may be specified in terms of the percent of words understood, percent of messages understood, percent of listeners that understand the set of messages, or a number of other similar parameters.

7.1 Access Phase

The parameters, access time, incorrect access probability, access denial probability, and access failure due to user blocking have been previously defined. Some values, or ranges of values pertinent to the system under consideration, and some classes of users are discussed below.

It should be noted that the AUTOVON is a dynamic system in that it is continuously being changed to adjust to the traffic patterns.

7.1.1 Access Time

For the AUTOVON system the average times given in Table 46 are representative, but do not necessarily apply to any specific area.

Table 46. Average Times (seconds)

	Within Area	Between Area
Off Hook - Start Dial	2.5	2.5
Start - Complete Dial	6.0	7.5
End Dial - Start Ring	7.0	12.0
Ring - Answer	9.0	9.0
Total	24.5	31.0

Some values of these parameters are available for telephone circuits (Linfield, 1979). For an average telephone channel the average access time is 35 seconds, rotary dial, conventional signaling, 28 seconds for multiple frequency dial, conventional signaling, and 20 seconds for MF dial, CCIS signaling. The disengagement time is 4 seconds on the average. Duffy and Mercer, (1978) break these numbers into a number of different classes such as off-hook to start of dialing, start of dialing to end of dialing, end of dialing to ring before an answer or a disconnect, end of dialing to answer without a ring signal, start of ringing to answer, and starter ringing to disconnect without an answer, among others. They give the means, standard deviations, and also cumulative distribution percentage points for each of these activities. In addition to successful completion they give times for call setup in the environment when the call was unsuccessful. There is no parameter in the digital parameter set which describes this.

Using the Harris Modem (Perkins and McRae, 1978) would increase these times by from 8 seconds to 15 seconds because of the synchronization time involved.

Crypto devices may also require additional time for setup.

7.1.2 Access Failure

Before discussing the failures, it is interesting to compare measurements of the access success rate for the commercial DDD network (Duffy and Mercer, 1978) with measurements made on the European AUTOVON system (GTE, 1977). For DDD, the access success rate was 69.8% for calling distances between 26 and 400 miles (41 and 643 km) and 72% for distances greater than 400 miles. For the European AUTOVON, the overall success rate was 47% for busy hours.

For the AUTOVON system there are usually two blocked access probabilities measured, access from the PBX to the AUTOVON switch, and the probability of being unsuccessful if access is obtained. The GTE (1977) study reports a probability of blocked access to the AUTOVON of 17% for busy hours. Given that access is achieved, the probability of being unsuccessful in completing the call is given as 46%, most of which is attributed to the unavailability of trunk lines out of the AUTOVON to the PBX's.

The Harris Modem achieved synchronization success in an average of 96.4% of the cases over a number of different AUTOVON configurations. This would imply an access denial probability increase of approximately 3.6%.

Similarly, increases in access denial would be expected with the addition of crypto gear, but would be expected to be small.

7.2 Transfer Phase

Some of the transfer phase parameters defined for the digital service mode are applicable to the analog service using digital transmission. However, application is in the form of a diagnostic which indicates those system characteristics that contribute to the three parameters: intelligibility, acceptability, and speaker recognition. In the following subsections, some of these relationships are exhibited.

7.2.1 Delay

We illustrate in this subsection an example relating the digital parameters to the analog parameters.

Some information is available on the listener acceptability of systems which have delays comparable to those for one or more satellite hops. Opinion Research Corporation, (1975) gives figures of 70% acceptable for one satellite hop, dropping to 40% for two hops, and 39% for three hops (corresponding to roughly 1.5 s round trip delay for the three hops). Kirkland and McDonald (1979) present a graph showing that 10% of the people rated circuits unacceptable with no delay, and 20% of the people rated the connection unacceptable with a delay corresponding to the U. S. Europe satellite circuit which is approximately 525 milliseconds round trip. These circuits were equipped with standard echo suppressors. No data are presently available on the acceptability of delays corresponding to one or more satellite hops for users such as some of the military personnel who might be using these circuits on a continuous basis. Experience using the circuit might lead to a much higher rate of acceptability of those circuits which are digital, without the noise introduced by the additional echo suppressors.

These delays, for digital transmission of voice, are comparable to the block transfer time. However, for voice systems, variability in the block transfer time must be kept small or the system will become unacceptable to the user. In addition, variable delays can cause the modem to lose synchronization resulting in loss of intelligibility or dropouts.

One analog parameter related to the delay that is usually mentioned in the context of voice circuits is echo. For the 16 Kb/s CVSD system or for most systems that use digital transmission, echo is not the same problem that it is for analog circuits since the voice cannot add in at the delay time unless it is transmitted some distance in an analog form which is not considered to be the case here. Echo contributes to the (un)acceptability of a system.

7.2.2 Values for the Analog Service Parameters

A number of measurements are presented in the literature which are applicable to CVSD, AUTOVON, the Harris Modem, or possible military user groups. Here, CVSD is used in the generic sense with the given values applying to the class of CVSD systems, and not specifically to the system for use over the AUTOVON network.

For the intelligibility objective parameters, the normalized energy and the band-weighted S/N, the CVSD system produces values from 92% for a zero error rate to 80% for 5% bit error rate. These parameters correspond to isolated word intelligibility testing and are representative of the average (over the listeners) of the percent of words understood correctly. No information is available about the distribution for each listener over the words or the distributions of the variability over the listeners.

To relate these numbers to performance over the AUTOVON net, tests using the Harris modem produced the distribution of errors given in Table 47. Note

Table 47. Summary of 16 kb Test Results[1]

Type of Call	# of Calls	Median BER	% of Looped Calls with BER ≤ 5%	2%	1%	% of One-Way Calls with BER ≤ 5%	2%	1%	Synchronozation Performance Tries	Successes	%
European IST Loops*	85	1.03E-3	92	80	70	100	98	96	346	330	95.4
Pacific IST Loops*	25	4.50E-4	100	99	97	100	100	100	93	93	100
CONUS IST Loops*	78	4.70E-3	100	85	80	100	100	100	234	234	100
European IST One-Way	21	5.40E-4	-	-	-	100	96	90	91	87	95.6
Trans-Atlantic One-Way	32	4.00E-3	-	-	-	100	92	75	161	161	100
Trans-Pacific One-Way	24	1.18E-4	-	-	-	100	100	99	109	109	100
European Access Loops	34	3.80E-5	100	99	97	100	100	100	151	139	92.1
European Remote Access Loops	26	1.03E-2	85	66	46	95	93	88	118	102	86.4
Pacific Remote Access Loops	9	3.80E-3	89	73	67	100	100	100	27	27	100
Totals	334								1330	1282	96.4

[1]Perkins and McRae (1978). *IST Interswitch trunks.

that the unsuccessful synchronization attempts would contribute to the access denial statistics.

Measurements of the acceptability of a 16 Kb/s CVSD system are reported by Voiers (1976). These measurements represent a rating of acceptability assigned by a listener panel consisting of 90 potential users of the military system. The listeners were instructed to score the system and system condition on a scale of 0 to 100%, with instructions that a normal telephone channel would score 90%. The scores for the 16 Kb/s CVSD were 60.7% for zero errors and 46.1% for the 5% bit error rate. For comparison purposes a clear 4 kHz channel was scored 88.8% by the panel. (This channel would be superior to the telephone channel for which the suggested score was 90%.) No statistical distributions over the listeners are available for this scoring, and no information is available for other user groups. It was reported that no generals and no clerical personnel were involved in the scoring.

For speaker recognition no scores have been reported for the 16 Kb/s system. Several investigators reported their judgement was that this system had good speaker recognizability properties. For the 5% bit error rate it has also been reported that the speaker recognizability properties are equivalent to a 9.6 Kb/s CVSD system with no bit error rates. These values are the opinions of a few experimenters.

Tables 48 and 49 summarize the available values for the objective measures and for the subjective interpretations. It is clear that many additional values need to be obtained to complete filling in the blanks, to get additional information about distributions of these numbers, and to specify numbers for appropriate user groups.

7.3 Summary of Analog Service Parameter Values

In this section we have reported values of the parameters pertinent to voice performance over the AUTOVON system using a 16 Kb/s CVSD with a Harris Modem. Three types of values are reported here:

1. system oriented values which are pertinent to the AUTOVON, the CVSD, or the Harris Modem;

2. user oriented values for different systems and different user groups, and;

3. a few user oriented values which represent one possible user group response to the CVSD system.

Table 48. Parameter Values for a 16 Kb/s CVSD with Zero Errors

Objective Measure \ Subjective Interpretation	Intelligibility		Acceptance		Speaker Recognition	
	(1)	(2)	(1)	(2)	(1)	(2)
Normalized Energy	~92%*	~92%	92%	60.7%***	---	"good"
Log Area Ratio	---	"	---	"	---	"
Short Term S/N	---	"	---	"	---	"
Band Weighted S/N	~65%**	"	65%	"	---	"
Speaker Recognition	---	"	---	"	---	"

(1) Objective Measure
(2) Listener Score
--- No Numbers Available
* Gamauf and Hartman (1977)
** Steeneken and Houtgast (1979)
*** Voiers (1976)

Table 49. Parameter Values for a 16 Kb/s CVSD with 5% BER

Subjective Interpretation / Objective Measure	Intelligibility		Acceptance		Speaker Recognition	
	(1)	(2)	(1)	(2)	(1)	(2)
Normalized Energy	~80%*	~80%	~80%	~46.1%***	---	---
Log Area Ratio	---	"	---	"	---	---
Short Term S/N	---	"	---	"	---	---
Band Weighted S/N	~54%**	"	~54%	"	---	---
Speaker Recognition	---	"	---	"	---	---

(1) Objective Measure
(2) Listener Score
--- No Numbers Available
* Gamauf and Hartman (1977)
** Steeneken and Houtgast (1979)
*** Voiers (1976)

The values for the access phase reported here are values for the system. At present, these values are changed for different user groups by establishing priority classes. The user values for these classes are unknown. It is assumed that the user groups were defined and some user values were established before assigning the priorities. In particular, the access denial probability is much smaller for high priority users than the values reported here for the system. The availability of numbers for this phase is summarized in rows 1, 2, and 3 of Tables 50 and 51. It should be noted that the military user groups for which data are available (Table 51) may not be appropriate for some other military group.

For the transfer phase, values which are directly applicable to military users or to the specific system are scarce. In the case of several parameters such as log-area ratios, values have been measured for comparison with the subjective scoring, but the values are not reported. For the delay parameter, delays are reported through certain portions of the system, but values are not given for round trip delays.

However, for a number of other systems, both user oriented and system oriented values of parameters are reported. This extensive background of data should assist in the assignment of the user values when the user groups are defined. Table 52 gives some typical values of these parameters.

8. OVERALL APPROACHES AND UNRESOLVED ISSUES

It is useful to place the work described in this report into its proper perspective. By doing so, some of the remaining issues come to light and the directions for future work in this 'user requirements' area become more clear. That is the purpose of this section.

New military systems, including telecommunication systems normally evolve through a number of sequences or phases. Five such phases typically used for new system developments are:

1. Concept Definition
2. System Validation
3. System Development
4. Design and Production
5. Implementation and Operation

The concept development phase itself may require a number of steps. The first step involves the specification of user requirements based on mission requirements.

Table 50. Availability of System Oriented Values for Analog Service Performance Parameters

		A	B
	Part A - Primary Parameters		
1.	Access Time	S	E
2.	Incorrect Access Probability	N	S
3.	Access Denial Probability	S	E
4.	Intelligibility		
	(a) Normalized Energy	S	E
	(b) Log Area Ratios	U	U
	(c) Short-Term S/N	U	U
	(d) Band-weighted S/N	S	E
5.	Acceptability		
	(a) Normalized Energy	S	E
	(b) Log Area Ratios	U	U
	(c) Short-Term S/N	U	U
	(d) Band-weighted S/N	S	E
6.	Speaker Recognition		
	(e) Computer Speaker Recognition	N	S
7.	Delay		
	(f) Round Trip Delay	U	S
8.	Disengagement Time		U
9.	Disengagement Denial Probability		U
	Part B - Secondary Parameters		
10.	Service Time Between Outages		U
11.	Outage Duration		U
12.	Outage Probability		U
	Part C - Ancillary Parameters		
13.	User Access Time Fraction		U
14.	User Delay Time Fraction		U
15.	User Disengagement Time Fraction		U

A = CVSD/AUTOVON System
B = Other Systems

N = None; S = Some; E = Extensive; U = Measured but Unavailable

Table 51. Availability of User Oriented Values for Analog Service Performance Parameters

		A	B
	Part A - Primary Parameters		
1.	Access Time	U	S
2.	Incorrect Access Probability	N	S
3.	Access Denial Probability	U	S
4.	Intelligibility	S	E
	(a) Normalized Energy		
	(b) Log Area Ratios		
	(c) Short-Term S/N		
	(d) Band-weighted S/N		
5.	Acceptability	S	S
	(a) Normalized Energy		
	(b) Log Area Ratios		
	(c) Short-Term S/N		
	(d) Band-weighted S/N		
6.	Speaker Recognition	N	S
	(e) Computer Speaker Recognition		
7.	Delay	N	S
	(f) Round Trip Delay		
8.	Disengagement Time		U
9.	Disengagement Denial Probability		U
	Part B - Secondary Parameters		
10.	Service Time Between Outages		U
11.	Outage Duration		U
12.	Outage Probability		U
	Part C - Ancillary Parameters		
13.	User Access Time Fraction		U
14.	User Delay Time Fraction		U
15.	User Disengagement Time Fraction		U

A = Military Users
B = Other Users

N = None; S = Some; E = Extensive; U = Measured but Unavailable

Table 52. Average System Oriented Values for Analog Service Performance Parameters

		A	B
	Part A - Primary Parameters		
1.	Access Time		33.2 s
2.	Incorrect Access Probability		1.7%[2]
3.	Access Denial Probability	54%[1]	29%[2]
4.	Intelligibility		
	(a) Normalized Energy		
	(b) Log Area Ratios		
	(c) Short-Term S/N		
	(d) Band-weighted S/N		
5.	Acceptability		
	(a) Normalized Energy		
	(b) Log Area Ratios		
	(c) Short-Term S/N		
	(d) Band-weighted S/N		
6.	Speaker Recognition		
	(e) Computer Speaker Recognition		
7.	Delay		
	(f) Round Trip Delay		11.7 ms[3] 37.3 ms[4]
8.	Disengagement Time		
9.	Disengagement Denial Probability		
	Part B - Secondary Parameters		
10.	Service Time Between Outages		
11.	Outage Duration		
12.	Outage Probability		
	Part C - Ancillary Parameters		
13.	User Access Time Fraction		
14.	User Delay Time Fraction		
15.	User Disengagement Time Fraction		

A = European AUTOVON
B = DDD Network

1 = Busy Hours; 2 = Busy Three Hours; 3 = 180-360 Miles (289-579 km); 4 = 1450-2900 Hours

Based on these user requirements, a number of system alternatives would be evaluated and functional designs developed for the preferred alternative. Engineering models are then constructed for testing to determine technical feasibility and to evaluate whether a given concept meets requirements.

Throughout these concept development steps and into the subsequent validation and system development phases, a continuing iterative process is required to specify the values for performance parameters in the final design and production phase. This iterative process is illustrated in Figure 61.

The iterations begin with a selected set of performance parameter values for a desired service. During the concept development phase initial values are assigned to the set based on user and mission requirements and an existing data based derived from past measurements. These initial values provide design goals for the initial system design for the validation phase. This initial system design also provides a base for estimating system costs. A cost-benefit analysis can then be performed. This may lead to a reassignment of values to the users' performance parameters. This process is then repeated by redesign and reanalysis in the system development phase and ultimately results in a final specification for production systems.

For upgrading existing systems the iterative process is somewhat different. The initial values are specifed for the parameters selected on the basis of new mission requirements. These values can be compared with measured performance on the existing system. If the existing system does not meet these new requirements, then design modifications, cost estimates, and cost-benefit analyses are undertaken to finalize the specifications for the final system modifications.

This report has sought to establish user-oriented performance parameters for application to the DCS II. We have been primarily considering the initial concept development phase. One requisite was that the performance parameters selected be simple, complete, broadly applicable, and easily measured. We have found that this requisite is by no means easily realized because the selection and structuring of the parameter set itself involves users and their perception of what they need. The diversity of user types and service classes makes this a complex task.

The assignment of initial values to the parameter set has also proved to be a challenge. Quantitive relationships between user-required and system-offered parameter value sets are not known. The assignment of numerical values can seldom be done independently of other parameters, or without reference to some part of the system. The difficulty is more than a lack of past measurement data. There also appears to be a general state of incomplete knowledge about the benefits of enhancing services, the basic or absolutely lowest acceptable user needs, and the cost that meeting these minimum needs would entail.

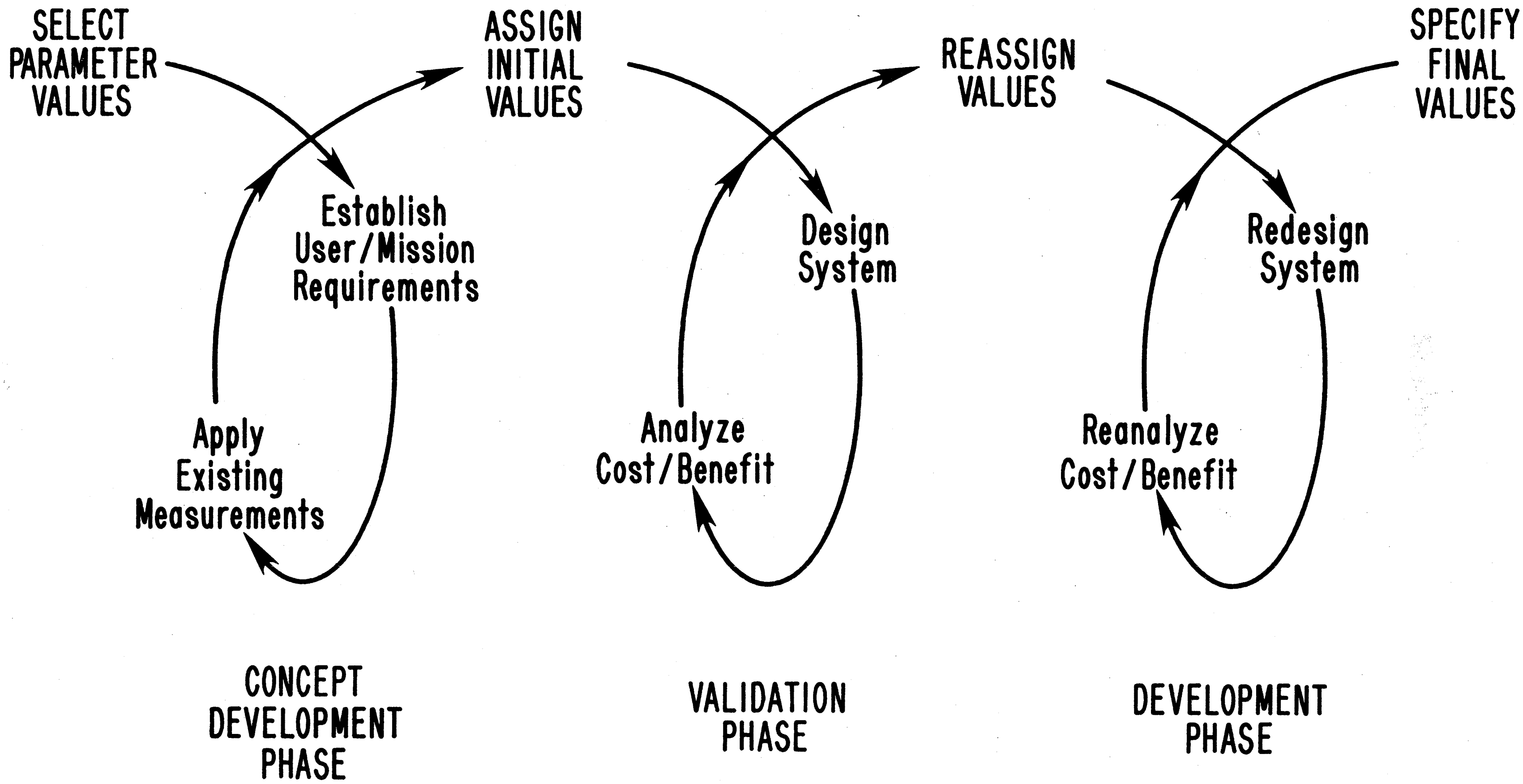


Figure 61. Iterative process for assigning values to user-oriented performance parameters.

Despite this incomplete state of knowledge, we have attempted to assign values to user-oriented service parameters for a digital service mode and an analog service mode in Sections 5 and 7 of this report. Project schedules and funding constraints have imposed limiting conditions on the scope and techniques employed. Only two modes of operation have been scrutinized and only the first iteration for assigning initial values has been attempted. No cost/benefit analysis has been performed.

Faced with extremely limited user service data for the DCS our approach has been to utilize everything pertinent that could be found. This included technical literature, engineering estimates, and a common sense inference of what the users and systems may realistically be in the 1980's.

The next step in this iterative process, and one to be considered for future work in this area, involves the development of a much firmer data base. Two approaches are of interest which were mentioned previously in Sections 5 and 6. One approach leads to the cost/benefit analysis required in subsequent system development phases. We expand on these approaches here noting that the more optimum or ideal the approach, the greater knowledge of user requirements knowledge it seems to require.

For lack of a better name, the first approach is called "the parametric effectiveness method." In this method the performance parameter variations are related to specific effects that they have on most, if not all, of the user groups.

The specific effects are made quantitative in terms of such tradeoffs as increased user productivity, mission speed, reduction of waste, and so forth. The overall benefit is often thought of as a generalized dollar net gain, that counteracts the costs associated with the system. The magnitude of the benefit quantity depends on the services that the system provides, and on the nature of the mission that the user faces. The parametric effectiveness method is quite general. Its major shortcoming is the obscure nature of the mapping function between performance parameters and dollar benefits.

The second method may be denoted as the "user opinion approach." The users are provided with services of certain quality and are asked to rate them. Thus, in this method objective performance measures are correlated with subjective assessments by specific user groups. The correlation can be accomplished in several ways, such as interviews, opinion polling, controlled laboratory testing, and so forth. Some of these tests were reviewed in Section 6. The user opinion approach works best when all the subjective evaluations involve users with a common mission and common service type. Otherwise, the user opinion approach suffers from poor repeatability and results that appear hard to interpret.

Usually the parametric effectiveness and the user opinion approaches yield ranges of values for performance parameters and ranges of effectiveness or perceived scores. Both depend further on system classes and user service groups. In order to assign single useful numbers to each parameter for system design, several additional steps are required. One key step involves the previously mentioned system cost evaluation. If both user effectiveness and system cost are functions of performance parameter values, then it should be possible to select optimum values on a cost-effective basis. The selection should be centered in the region where the total user benefit exceeds the cost outlay by the biggest margin.

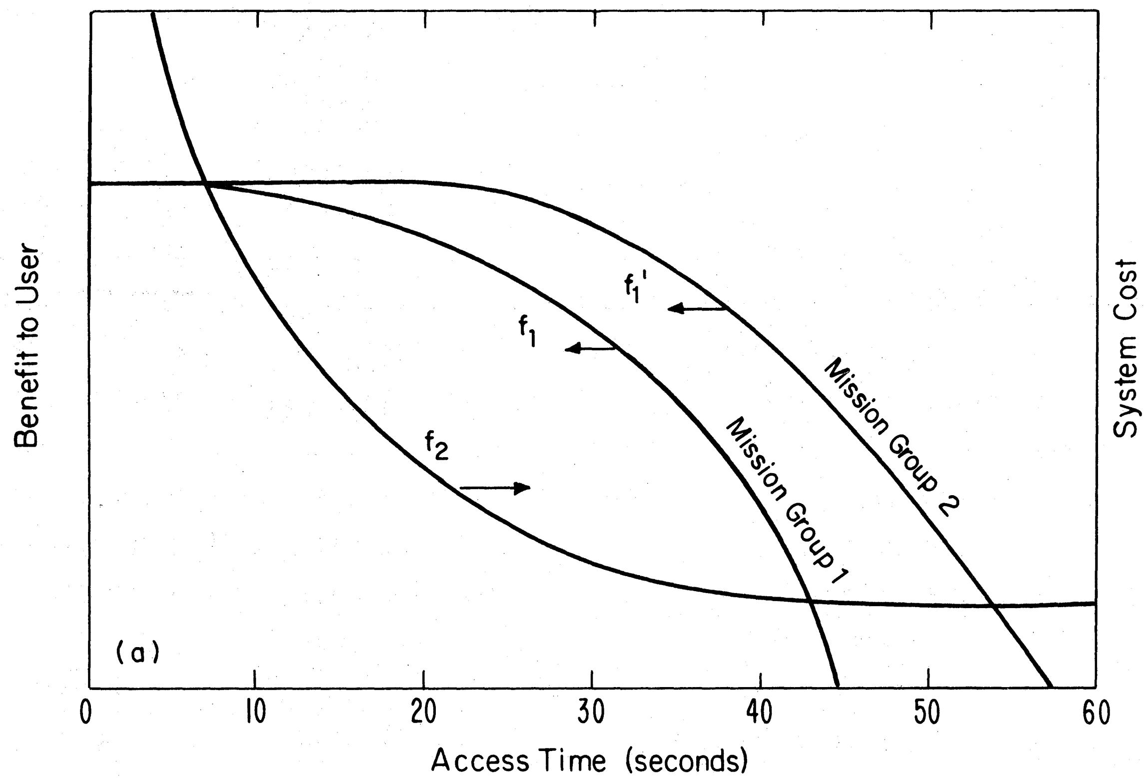
These two approaches can be illustrated by examples using quite qualitative estimates for the numbers involved.

The first example, a parametric effectiveness approach, is illustrated in Figure 62(a). It depicts service access time on a voice network as the performance parameter of interest. As noted in Sections 4 to 7, access times are concerned only with successful access attempts. Access denial and incorrect access probabilities are handled as separate performance parameters.

The benefit to the user in this illustration is shown to increase with decreasing access time. This is represented by curves marked f_1 and f_1' for two different user groups. As these benefits increase, the cost of achieving the desired performance with a given type system is also expected to increase. The latter fact is indicated in Figure 62(a) by the curve marked f_2. Quantitative costs are not shown, but must be deduced for both actual existing systems and potential future systems.

One knows, for instance, that a circuit switched network with per channel signaling and dc pulsing, plus a rotary dial, may permit user access times no better than 30 to 40 seconds. Access times in the 20-to-30 second range would require common channel interswitch signaling and possibly dual tone, multifrequency, push button, dialers. This involves a substantial increase in the system cost. Access times less than 10 seconds would require some form of high-speed dialing. Ultimately, for access times approaching zero, a dedicated nonswitched network would be needed with another substantial increase in cost.

Figure 62(b) illustrates (again qualitatively) one means of assigning a value for access time for each mission group. A measure of net value to the user is obtained by taking the difference between the curves f_1 (or f_1') and f_2 in Figure



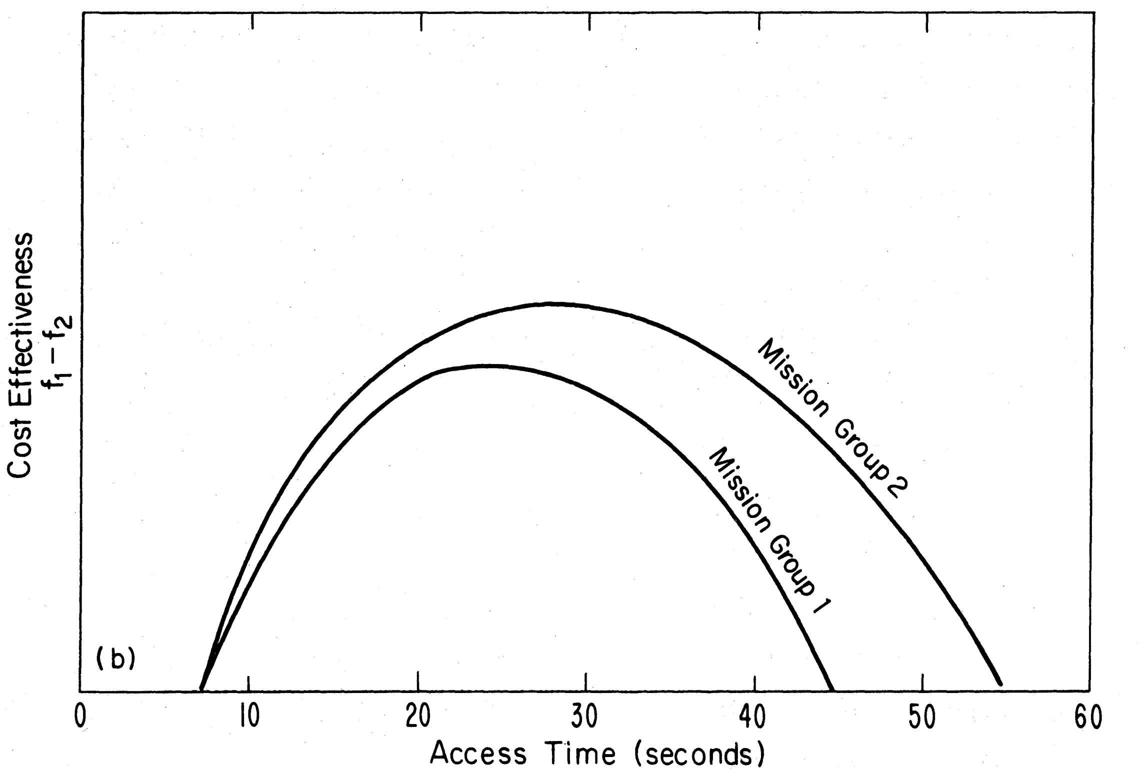


Figure 62. Functional analysis approach for assigning value to access time.

62(a). The curves in 62(b) provide a general result. Design values may be selected from the broad maximums indicated.

The previously introduced user opinion approach does not involve cost considerations. Here each mission group is alleged to know the largest access time permissible to accomplish its mission. A user poll is conducted to obtain the distribution of either dissatisfied or satisfied customers as a function of observed access time. Figure 63(a) shows the percentages of dissatisfied users as a function of access time. Two mission groups are depicted. For mission group 1, access time of 20 s causes 35% of the users to be dissatified; access time of 35 s leaves 95% dissatisfied. For mission group 2, only 5% are unhappy with an access time of 20 seconds. In principle, user-oriented access time values can be selected for each user group based on percentage goals of satisfied users.

Access time distributions can also be obtained by actual measurement on selected systems. Two such empirical distributions are illustrated in Figure 63(b). One curve shows that system A meets the 20 s access time objective, or better, in 95% of the trials. System B meets the same requirement about 25% of the time. A composite view from parts (a) and (b) of Figure 63 is also useful. It reveals that system choice must take into account requirements of all mission groups to be served.

In this report we have followed, where possible, the user opinion approach for the first interation toward assigning values to performance parameters. However, there were too many instances where no numerical user opinion data were available. In such cases, values were assigned on the basis of what was judged achievable with reasonable present or near future technology. In some instances, extrapolation of both user opinions and system capabilities was necessary because of the limited general knowledge of performance requirements.

The parametric effectiveness method is ideally suited for system validation and system development phases where system costs can be estimated with reasonable confidence.

9. ACKNOWLEDGMENTS

This report is a first attempt at describing communication network performance from a user's viewpoint and applying these results to specific systems. The approach is new in many respects, and the results presented here are far from definitive. Much remains to be done, particularly in the area of user parametric effectiveness.

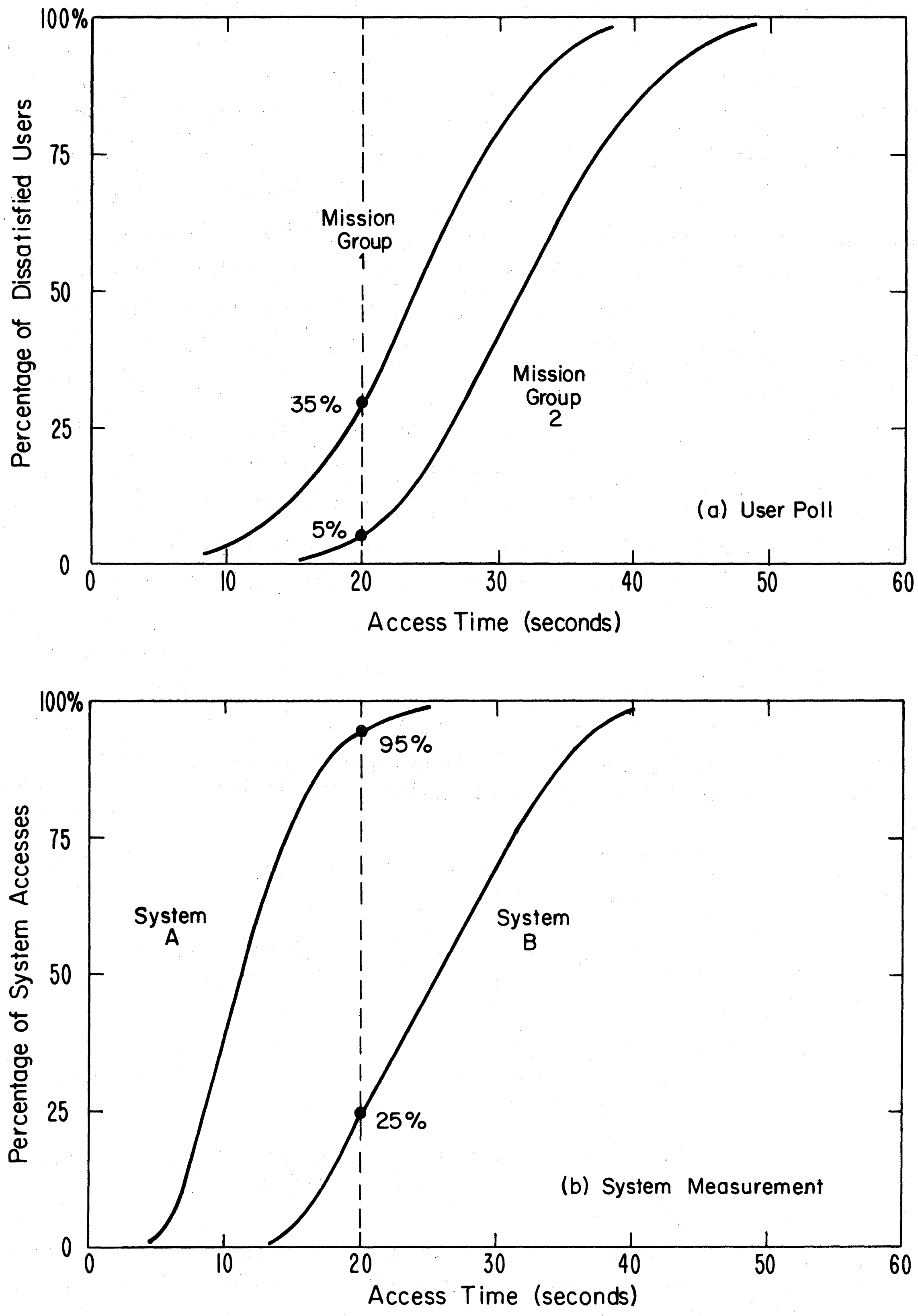


Figure 63. User opinion approach for access time selection.

A number of persons, in addition to the authors, contributed to this work. Mr. N. B. Seitz, who co-authored one of the appendices, also provided inspiration and made pertinent suggestions which have been incorporated. His previous work in the development of the user-oriented data communications standard FS-1033 provided a valuable data base to follow. Dr. P. M. McManamon, Associate Director of ITS for the Advanced Communication Networks Division participated in many of the technical discussions and was instrumental in resolving technical issues and suggesting approaches to be taken. Finally Mr. M. Horowitz of DCEC, the contract monitor, provided useful insight into the military user's needs and the DCS II system requirements. His persistence led to some important reconsiderations in selecting modes of operation, defining performance parameters, and in selecting values for the parameters.

In any new approach, like the one presented here, there are a number of diverse questions to be resolved. The authors themselves often could not reach a consensus on all issues. Some of these differences in opinion are reflected in the report. We hope that the readers will recognize these differences as they arise and, in turn, will assess them with their own application or interest in mind.

10. REFERENCES

Abrams, M.D. (1974), A new approach to performance evaluation of computer networks, Proc. IEEE/NBS Symp. Comput. Networks: Trends and Applications, Gaithersburg, MD, 15-20, May.

Abramson, N. (1977), The throughput of packet broadcasting channels, IEEE Trans. on Comm., COM-25, 117-128.

Agrawal, A., and Wen C. Lin (1974), An on-line speech intelligibility measurement system, IEEE Trans. on Acoustics, Speech and Signal Processing, ASSP-22, 203-206.

ANSI (1971), Proposed American national standard and determination of performance of data communication systems, X3S3.5/80, American National Standards Institute, 1430 Broadway, New York, NY.

ANSI (1980), Proposed ANSI standard data communication performance parameters (Revision 80.10), X3S35/125, American National Standards Institute, 1430 Broadway, New York, NY.

Atal, B.S. (1974), Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification, J. Acoust. Soc. Amer., 55, 1304-1313.

Atal, B.S. (1976), Automatic recognition of speakers from their voices, Proc. IEEE, 64, 460-475.

AT&T (1970), Data communications using switched telecommunications network, Bell System Technical Reference (American Telephone and Telegraph Co., New York, NY).

AT&T (1977), Telecommunications transmission engineering (American Telephone and Telegraph Co., New York, NY).

Audin, G. (1978), Performance: the battle of needs vs. desires, Data Communications, 7, 99-105.

Barnwell, T.P., and W.D. Voiers (1979), An analysis of objective measures for user acceptance of voice communications systems, Final Report DA100-78-C-0003, DCEC, Reston, VA 22090, September.

Benice, R.J., and A.H. Frey (1964a), An analysis of retransmission systems, IEEE Trans. Commun. Technol., COM-12, 135-145.

Benice, R.J., and A.H. Frey (1964b), Comparisons of error control techniques, IEEE Trans. Commun. Technol., COM-12, 146-154.

Berlekamp, E.R. (1968), Algebraic Coding Theory (McGraw-Hill, New York, NY).

Boies, S.J. (1974), User behavior on an interactive computer system, IBM System Journal, No. 1, 2-18.

Brayer, K. (1978a), Time, not throughput, for net designs. Part 1: techniques, Data Communications, 7, 39-46.

Brayer, K. (1978b), Time, not throughput, for net designs. Part 2: application, Data Communications, 7, 83-92.

Buhrke, R.E., and J.J. Mele (1974), Field tests of the first and second GTE Automatic Electric TSPS sites, GTE AE Techn. Journal, 14, 176-183.

Cole, G.D. (1972), Performance measurements on the ARPA computer network, IEEE Trans. on Comm., COM-20, 630-636.

Coviello, G., and P.A. Vena (1975), Integration of circuit/packet switching by a SENET (slotted envelope network) concept, Proc. of NTC '75, New Orleans, LA, 42-12 to 42-17.

Crochiere, R.E., L.R. Rabiner, S.S. Jayant, and J.M. Tribolet (1978), A study of objective measures for speech waveform coders, International Zurich Seminar on Digital Communications, Zurich, Switzerland, March.

Crow, E.L., and M.J. Miles (1977), Confidence limits for digital error rates from dependent transmissions, OT Report 77-118, National Telecommunications and Information Administration (Available from NTIS, 5285 Port Royal Road, Springfield, VA 22161, Accession No. PB 267 777).

Crow, E.L. (1979), Statistical methods for estimating time and rate parameters of digital communication systems, NTIA Report 79-21, Boulder, CO, June.

Daumer, W.R., and J.R. Cananaugh (1978), A subjective comparison of selected digital codecs for speech, Bell Syst. Tech. J., 57, 3119-3165.

DCA (1975), System Performance Specification for AUTODIN II Phase I, Defense Communication Agency, Reston, VA 22090, November.

Duffy, F.P., and R.A. Mercer (1978), A study of network performance and customer behavior during direct-distance-dialing call attempts in the U.S.A., Bell Syst. Tech. J., 57, 1-33.

Easton, M.C. (1980), Batch throughput efficiency of ADCCP/HDLC/SDLC selective reject protocols, IEEE Trans. Comm., COM-28, 187-195.

Elam, P.G. (1978), Considering human needs can boost network efficiency, Data Communications, 7, 51-60.

Fairbanks, G. (1958), Test of phenomic differentiation: the rhyme test, J. Acoust. Soc. Am., 30, 596-600.

Feldman, N.E., W. Sollfrey, S. Katz, and S.J. Dudzinsky, Jr. (1979), Writer-to-reader delays in military communications systems, RAND Corp. Report R-2473-AF, Santa Monica, CA, October.

Flanagan, J.L., M.R. Schroeder, B.S. Atal, R.E. Crochiere, N.S. Jayant, and J.M. Tribolet (1979), Speech coding, IEEE Trans. Comm., COM-27, 710-737.

Frank, H., and L. Hopewell (1974), Network reliability, Datamation, 85-91, August.

Frank, H., I. Gitman, and R. Van Slyke (1976), Local, regional and large scale integrated networks, 5 Volumes, Sixth Semiannual Technical Report, Network Analysis Corporation, Glen Cove, NY 11542, February.

FTSC (1973), Joint Federal Telecommunication and Information Processing Standard for "Synchronous high speed signaling rates between data terminal equipment and data communication equipment," GSA, Washington, DC.

Gamauf, K.J., and W.J. Hartman (1977), Objective measurement of voice channel intelligibility, FAA Report No. FAA-RD-77-153, Washington, DC 20590, October.

Gierhart, G.D., R.W. Hubbard, and D.V. Glen (1970), Electrospace planning and engineering for the air traffic environment, Federal Aviation Administration, Systems Research and Development Service, Washington, DC 20590, Report No. FAA-RD-70-71, December.

Gitman, I., and H. Frank (1978), Economic analysis of integrated voice and data networks: A case study, Proc, IEEE, 66, 1549-1570.

Gray, A.H., Jr., and J.D. Markel (1976), Distance measures for speech processing, IEEE Trans. on Acoustics, Speech, and Signal Processing, ASSP-24, 380-391.

Green, P.E. (1979), An introduction to network architectures and protocols, IBM Systems Journal, 18, No. 2, 202-222.

Griffiths, J.D. (1967), Rhyming minimal contrasts; a simplified diagnostic articulation test, J. Acoust. Soc. Am., 42, 236-241.

Grubb, D.S., and I.W. Cotton (1978), Rating performance, Executive Guide to Data Communications (McGraw-Hill, New York, NY), 57-63.

GSA (1979), Interim Federal Standard Telecommunications: Digital communication performance parameters, INT-FED-STD-001033, Washington, DC, August.

GTE Sylvania Incorporated (1977), Traffic flow, performance and interface characterization study in Europe, Final Technical Report, Needham, MA 02194, October.

Hanson, E.D. (1971), Intelligibility testing of the continuous variable slope delta (CVSD) coder-decoder (codec), ECOM R and D Technical Report, Ft. Monmouth, NT, February.

House, A.S., C.E. Williams, M.H.L. Hecker, and K.D. Kryter (1965), Articulation-testing methods: consonatal differentiation with a closed-response set, J. Acoust. Soc. Am., 37, 158-166.

Hubbard, R.W., and D.V. Glen (1968), Required frequency protection ratios for an FAA satellite air traffic control communication system, Unpublished[1] ESSA Tech. Memo. ERLTM-ITS-140[2], FAA Contract No. FA 67-WAI 130.

Hubbard, R.W., and W.J. Hartman (1974), Objective measurement techniques for evaluating voice communication channels, FAA Report No. FAA-RD-74-77, Washington, DC, July.

Kelley, K.G. (1977), An Evaluation of Data Information Transfer Requirements for the Future DCS, DCEC Technical Note 24-77, DCA, Reston, VA 22090, November.

Kimmett, F.G., and N.B. Seitz (1978), Digital communication performance parameters for proposed Federal Standard 1033, Volume II: Application examples, NTIA Report 78-4, Boulder, CO, May.

Kirk, K.W., and J.L. Osterholz (1976), DCS Digital Transmission System Performance, DCEC Tech. Report 12-76, DCA, Reston, VA 22090, November.

Kirkland, F.E., and K.M. McDonald (1979), Overseas communication by submarine cables, Telecommunications, 13 No. 5, May, 45-50.

Kleinrock, L. (1975), Queueing Systems, Volume I: Theory (John Wiley and Sons, New York, NY).

Kleinrock, L. (1976), Queueing Systems, Volume II: Computer Applications (John Wiley and Sons, New York, NY).

Kleinrock, L. (1978), Principles and lessons in packet communications, Proc. IEEE, 66, 1320-1329.

Kleinrock, L., W.E. Naylor, and H. Opderbeck (1976), A study of line overhead in the Arpanet, Comm. of ACM, 19, 3-13.

Kleinrock, L., and H. Opderbeck (1977), Throughput in the ARPANET - protocols and measurements, IEEE Trans. on Comm., COM-25, 95-104.

Kryter, K.D. (1962), Methods for the calculation and use of the articulation index, J. Acoust. Soc. of Am., 34, 1689-1697.

Kryter, K.D., and J.H. Ball (1964), A meter for measuring the performance of speech communication systems, Tech. Doc. Rept. No. ESD-TDR-64-674, Electronic Systems Division, Air Force Systems Command, L.G. Hanscom Field, Bedford, MA.

Kulkarni, V.R., P.J. Nichols, N.P. Schmitt, and R.E. Hunter (1979), AUTODIN II backbone delay analysis (Draft), Western Union System Eng. Technote 79-07, June.

Levine, R.H. (1976), The evolving DCS: Introduction and overview, ICC Conference Record, Vol. II, Philadelphia, PA, June, 33-2 to 33-4.

Licklider, J.C.R., and A. Vezza (1978), Applications of information networks, Proc. IEEE, 66, 1330-1346.

Lin, S., and P.S. Yu (1980), An effective error control scheme for satellite communications, IEEE Trans. Comm., COM-28, 395-401.

Linfield, R.F. (1979), Control signaling in a military switching environment, NTIA Report 79-13, Boulder, CO 80303, January.

Luck, J.E. (1969), Automatic speaker verification using cepstral measurements, J. Acoust. Soc. Amer., 46, 1026-1032, October.

Mahoney, J.J., Jr., J.J. Mansell, and R.C. Matlack (1975), Digital Data System: User's view of the network, Bell Syst. Tech. J., 54, 833-844.

Martin, J. (1972), Systems Analysis for Data Transmission (Prentice-Hall, Inc., Englewood Cliffs, NJ).

Martin, J. (1973), Design of Man-Computer Dialogues (Prentice-Hall, Inc., Englewood Cliffs, NJ).

McClary, W.S. (1977), AUTODIN II interface, AFCEA 2nd Annual Seminar, Ft. Monmouth, NJ, September.

McDermott, B.J., C. Scagliola, and D.J. Goodman (1978), Perceptual and objective evaluation of speech processed by adaptive differential PCM, Proc. of ICASSP, Tulsa, OK, April.

McRae, D.D., F.A. Perkins, D.J. Risavy, J.N. York (1976), 16 Kb/s data modem technologies, Technical Report TR-76-311 for the Rome Air Development Center, RADC Contract F 30602-75-C-0129, Harris Corporation, Melborne, Florida.

Miller, R.B. (1968), Response time in man-computer conversational transactions, Proc. of 1968 Fall Joint Computer Conference, 267-277.

Morris, J.M. (1978), On another Go-Back-N ARQ technique for high error rate conditions, IEEE Trans. Comm., COM-26 187-189.

Munson, W.A., and J.E. Karlin (1962), Isopreference method for evaluating speech-transmission circuits, The Journal of the Acoustical Society of America, 34, No. 6, 762-774, June.

O'Brien, P.J., and A.C. Busch (1969), Effects of selective system parameters on communications intelligibility, FAA Rept. No. AN-69-21 (RD-68-59), National Aviation Facilities Experimental Center, Altantic City, NJ 08405.

Occhiogrosso, L., L. Gitman, W. Hsieh, and H. Franks (1977), Performance analysis of integrated switching communications systems, Proc. of NTC '77, Los Angeles, CA, 12.4-1 to 12.4-8.

Opinion Research Corporation (1975), Perceptions of telephone transmission quality on selected circuits, North Harrison Street, Princetown, NJ 08540, Survey Research conducted for FCC, contract FCC-1062 (23E25.15), December.

Perkins, F.A., and D.D. McRae (1978), 16 Kb/s modem evaluation, Final Technical Report RADC-TR-78-127 for the Rome Air Development Center, RARC Contract 30602-76-C-0460, Harris Corporation, Melborne, Florida.

Peterson, W.W., and E.J. Weldon (1972), Error-Correcting Codes, Second Edition (The MIT Press, Cambridge, MA).

Reed, M.A., and T.D. Smetanka (1977), How to determine message response time for satellites, Data Communications, 6, 42-47.

Roberts, L.G. (1978), The evolution of packet switching, Proc. IEEE, 66 1307-1313.

Rose, M.P., and J.P. O'Keefe (1980), User studies value-added-network response times, Data Communications, 9, No. 4, 83-90.

Rosenberg, A.E. (1973), Automatic speaker verification: a review, Proc. IEEE 64, 475-487.

Rosner, R.D. (1973), A digital data network concept for the defense communications system, Proc. of NTC '73, Atlanta, GA, 22c-1 to 22c-4.

Rudin, H. (1974), The design of digital communications systems: Can traffic theory help?, Proc. of 1974 International Zurich Seminar on Digital Communications, Zurich, Switzerland, March, D5(1-6).

Sambur, M.R., and N.S. Jayant (1976), LPC analysis/synthesis from speech inputs containing quantizing noise or additive white noise, IEEE Trans. on Acoustics, Speech, and Signal Processing, ASSP-24, No. 6, December.

Sanders, R. (1980), Effects of switching technologies on network delay, Data Communications, 9, No. 4, 109-118.

Sastry, A.F.K. (1975), Improving Automatic Repeat-Request (ARQ) performance on satellite channels under high error rate conditions, IEEE Trans. on Comm., COM-23, 436-439.

Schwartz, M., R.R. Boorstyn, and R.L. Pickholtz (1972), Terminal-oriented computer-communication networks, Proc. IEEE, 60, 1408-1423.

Sciulli, J.A. (1980), Satellite circuit simulator for satellite data transmission systems, Telecommunications, 14, No. 3, 58-59, 89.

Seitz, N.B., and P.M. McManamon (1978), Digital communication performance parameters for proposed federal standard 1033, Vol I: Standard parameters, NTIA Report 78-4 (Available from NTIS, 5285 Port Royal Road, Springfield, VA 22151, Accession No. PB 283-580/AS.

Seitz, N.B., and D. Bodson (1980), Data communication performance assessment, Telecommunications, 35-89, February.

Sen, T.K., and J.D. Carroll (1973), Signal-to-noise ratio as a predictor of speech transmission quality, IEEE Trans. on Audio and Electroacoustics, 384-387.

Sergeant, L., J.E. Atkinson, and P.G. Lacroix (1979), The NSMRL tri-word test of intelligibility (TTI), J. Acoust. Soc. Am, 65, 218-222.

Sevcik, P.J. (1977), AUTODIN II subscriber access protocols and interfaces, Proc. of NTC '77, Los Angeles, CA, 37:6-1 to 37:6-6.

Shah, A.N. (1977), Impact of military requirements on packet switching network design, Proc. of NEC, Chicago, IL, 31, 218-221.

Shanmugam, K.S., and P. Balaban (1979), Estimation of error probabilities in a digital communication system, Proc. of NTC '79, Washington, DC, 35.5.1 to 35.5.5.

Stavroulakis, P. (1972), Analysis of new telecommunication services via system attributes, Bell Laboratories, Whippany, NJ.

Steeneken, H.J.M., and T. Houtgast, (1979), A physical method for measuring speech-transmission quality, Institute for Perception, Soesterberg, The Netherlands, Report No. IZF 1979-1.

Stewart, H.M. (1978), Performance analysis of complex communications systems, IBM Systems Journal, 18, No. 3, 356-373.

Streeter, D.N. (1972), Cost-benefit evaluation in scientific computer services, IBM Systems Journal, No. 3, 219-233.

Streeter, D.N. (1974), The Scientific Process and the Computer (John Wiley and Sons, New York, NY), Section IV.

Tribolet, J.M., P. Noll, B.J. McDermoth, and R.E. Crochiere (1979), A comparison of the performance of four low-bit-rate speech waveform coders, Bell Syst. Tech. J., 58, 699-712.

Vigants, A., and M.V. Pursley (1979), Transmission unavailability of frequency-diversity protected microwave FM radio systems caused by multipath fading, Bell Syst. Tech. J., 58, 1779-1796.

Voiers, W.D. (1967), Performance evaluation of speech processing devices, III. Diagnostic evaluation of speech intelligibility, Report No. SRRC-CR-67-6, Final Report, Sperry Rand Research Center, Sudbury, MA 01776.

Voiers, W.D. (1971), Present status of the diagnostic rhyme test, Presented at the meetings of the Groupement des Acousticiens de Langue Francaise, Groupe, "Communication Parlee," Aix-en-Provence, France.

Voiers, W.D. (1976), Methods of predicting user acceptance of voice communication systems, Final Report, Dynastat, Inc., 2704 Rio Grande, Austin, TX 78705, June.

Voiers, W.D. (1977), Diagnostic evaluation of speech intelligibility, Dynastat Inc., Austin, Texas 78705. To appear in Speech Intelligibility and Recognition, M.E. Hawley, (ed.) Dowden, Hutchinson and Ross, Stroudsburg, PA.

Voiers, W.D., M.F. Cohen, and J. Mickunas (1965), Evaluation of speech processing devices, I. Intelligibility, Quality, Speaker Recognizability, AFCRL-65-926.

Wagner, L.H. (1977), Access area digital switching system, Proceedings of Second Annual Seminar, AFCEA 2nd Annual Seminar, Ft. Monmouth, NJ, 39-57.

Webster (1960), Webster's New World Dictionary, (The World Publishing Co., New York, NY).

Wong, D.Y., and J.D. Markel (1978), An intelligibility evaluation of several linear prediction vocoder modifications, IEEE Trans. on Acoustics, Speech, and Signal Processing, ASSP-26, 424-435.

APPENDIX A: PROJECT OVERVIEW FOR DETERMINATION OF DIGITAL TRANSMISSION TECHNICAL CRITERIA

A-1. Introduction

This project overview describes the technical approach the Institute for Telecommunication Sciences (ITS) proposed to follow in conducting an engineering development project, entitled " Determination of Digital Transmission Technical Criteria," for the Transmission Engineering Division of the Defense Communications Engineering Center (DCEC).

A-1.1 Background

As the primary engineering arm of the Defense Communications Agency, the Defense Communications Engineering Center (DCEC) has the major responsibility for developing and implementing the second-generation Defense Communications System (DCS II). This planned system, which will gradually supplant the existing AUTOVON, AUTODIN, and AUTOSEVOCOM systems during the 1980's, has been motivated by two fundamental changes in the military communications environment:

1. A substantial increase in the demand for high quality data and secure voice communication services, in recognition of the powerful "force multiplier" effect of communications (Babbitt, 1977).

2. Dramatic improvements in the technologies of digital transmission and network resource sharing, typified by the development of the Digital European Backbone (DEB) and the ARPANET.

As described in recent Defense Department planning documents (DCEC, 1979; DCA, 1978), the second-generation DCS will differ from the current system in three major respects:

1. The analog circuit switches and FDM transmission facilities currently supporting AUTOVON will be replaced by digital equipment.

2. The traditional AUTODIN store-and-forware message switching network will be augmented, and to a growing extent replaced, by AUTODIN II - a second-generation packet switching system based on the ARPANET.

3. The existing AUTOSEVOCOM system will be replaced, under the Secure Voice Improvement Program, with improved narrowband secure voice services provided by the new AUTOVON facilities. The number of voice subscribers receiving security protection will be significantly increased.

The planned DCS II network represents a direct application of the new technologies of digital transmission and resource sharing to the post-1985 needs of military communications users. In comparison with its predecessor system, DCS II offers the potential of substantially improved end-to-end performance; broader geographical and organizational coverage; and more flexible adaptation to growth and change.

In addition to these operational benefits, the planned DCS II network offers the potential of major cost savings. The superior reliability and maintainability of the digital DCS will translate into reduce training, operation, and maintenance costs; and its improved performance will encourage military users to employ the common-user network in preference to inefficient, costly dedicated services. As an example of the latter potential, a recent report to the Congress by the General Accounting Office noted a savings of $100 K per year from the conversion of a single Navy dedicated system to AUTODIN - a yearly savings in excess of 20% of the total cost of the procured service. The total DOD budget for dedicated communication services was $117 million in 1977 (GAO, 1977).

While the potential benefits of DCS II are substantial, their realization will not be easy. Its designers face major problems in determining user requirements for services; in evaluating the performance of candidate facilities; and, perhaps most importantly, in relating these two variables. The success of the DCS II system will depend to a large extent on how effectively these problems are addressed during the next two to three years.

A-1.2 Project Objectives

The project outlined in this overview will contribute to the solution of the problems noted above by developing precise technical performance criteria to be used in the specification of DCS II systems and subsystems. The overall project has been divided into three major phases, with the following specific objectives:

Phase A - User Criteria. Develop user-oriented performance criteria for two representative DCS II services: a digital voice communication service and an interactive data communication service.

Phase B - Technical Criteria. Translate the user-oriented performance criteria into corresponding technical (engineering-oriented) criteria on an end-to-end basis.

Phase C - Subsystem Criteria. Develop user-oriented and technical performance criteria for additional DCS II service modes. Allocate the technical criteria to subsystems within a defined DCS II global reference network.

Although all relevant performance criteria will be considered, the project will focus on criteria that have a major impact on the design and specification of transmission subsystems.

A-2. Project Overview

This section summarizes the overall criteria development problem and the ITS approach to its solution. The section is basically an interpretation of the original DCEC work statement (Phases A through C) in light of the criteria development work conducted at ITS during the past five years (Seitz and McManamon, 1978; Gamauf and Hartman, 1977). This ITS work was summarized in our earlier comments on the draft DCEC work statement (ITS, 1978).

A-2.1 Definitions

A common understanding of certain fundamental telecommunication terms is considered essential. These terms are defined below. Where possible, related terms are clustered to facilitate understanding. Terms underlined within a particular definition are defined elsewhere in the list.

Communication system - an association of physical and functional elements whose joint activities provide one or more communication services to a population of end users. The major subsystems within an end-to-end communication system are transmission subsystems, switches, and communication terminals. The communication system element that interfaces with the end user is the communication terminal or, in the case of computer-to-computer communications, the host computer access method.

End user - a human terminal operator, unattended device medium, or computer application program which receives a specified communication service from an end-to-end communication system. The end users function as the ultimate source and sink of transferred user information.

User-system interface - any functional or physical boundary separating an end user from an adjacent communication system. In the case where the end user is a human terminal operator or unattended device medium (e.g., punched cards), the user-system interface is between the operator or medium and the terminal. In the case where the end user is a computer application program, the user-system interface is between that program and the local telecommunications access method.

Application program - a computer program which performs data processing functions under the control of an operating system - e.g., a text editor or compiler. Application programs constitute one of three general types of end users in teleprocessing systems. Application programs typically obtain communication service via system calls to an access method within the same Host computer.

Access method - a computer program which controls the provision of communication service to a group of application programs within a teleprocessing computer. An example is the Virtual Telecommunications Access Method (VTAM) in IBM's Systems Network Architecture (SNA).

Communication service - a defined set of functions performed by a communication system for its end users. The most fundamental of these functions is the transfer of user information between a source and destination user.

Digital service - any communication service in which the signals crossing the user-system interfaces are intended to represent a limited number of discrete levels or states. The term "data communication service" is considered synonomous with "digital service" in this report.

Analog service - any communication service other than a digital service. Familiar examples of analog services are voice, facsimile, and television transmission.

User information - any signal or symbol input to a communication system by a source user for ultimate delivery to one or more destination users. User information is distinguished from overhead information by the fact that it is intended to cross both a source and destination user interface, and is not intended to change the communication system state. Typical representations of user information at the end user/communication system interface are (1) a telephone subscriber's voice, (2) the graphic ASCII characters typed and displayed at a data terminal; and (3) a binary output file produced by a teleprocessing application program.

Overhead information - any communicated signal or symbol which does not constitute user information. Overhead information is distinguished from user information by the fact that its communication is intended to change the state of some communication system element; it does not normally cross both user-system interfaces. Typical examples of overhead information are (1) the position of a telephone hookswitch, dial, or pushbutton; (2) ASCII control characters such as ENQ, ACK, and SYN; and (3) system call messages passed between an applications program and the local "access method" in a teleprocessing computer (e.g., the CONNECT system call in the ARPA network).

Descriptor (or parameter) - a statistical quantity whose numerical values characterize a particular aspect of communication system performance - e.g., Bit Error Probability. Different descriptors apply, in general, to analog and digital services. The terms descriptor and parameter are considered synonomous for the purpose of this project.

Criterion - a numerical quantity which represents the specified or measured value of a performance parameter (e.g., Bit Error Probability value of 10^{-6}). Different criteria apply, in general, to each distinct user class of service. The terms criterion and value are considered synonomous for the purpose of this project.

User-oriented - a performance parameter or value which characterizes an end-to-end communication service as seen by the end users. User-oriented parameters and values express fundamental user concerns about communication performance, and are always defined and measured in terms of events which are directly observable to the end users. An example of a user-oriented performance parameter applicable to voice services is intelligibility. The term "operational" is sometimes used as a synonym for "user-oriented."

Engineering-oriented - a performance parameter or value which characterizes a communication subsystem as seen by the designer or operator. Engineering-oriented parameters express technical performance characteristics which are essential to system design and operation, but are not, in general, relevant or observable to the end users. An example of an engineering-oriented parameter applicable to digital voice systems is channel signalling rate. The term "technical" is used interchangeably with "engineering-oriented" in this report.

User class - a distinct set of users having the same or similar service requirements. Typical military user classes are strategic, tactical, and administrative.

Service mode - a particular sequence of interactions by which a communication service is provided. A service mode is formally defined by a transaction profile. "Virtual circuit" and "datagram" are two distinct service modes provided in modern packet switching networks.

An additional concept involved which need definition is the concept of an aggregate user. Although the primary focus of Phase A is on characterizing end-to-end services, it will also be necessary to consider the input/output performance of subsystems terminated within the end user interfaces (e.g., the digital transmission backbone). In such situations, we regard the chain of elements on the drop side of each subsystem interface as an aggregate user of the subsystem; and apply the concepts of communication service, etc. at the aggregate user and subsystem interfaces. An aggregate user thus consists of an end user plus one or more adjacent system elements, which collectively receive communication service from a subsystem. The aggregate user concept facilitates the allocation of end-to-end performance objectives to subsystems.

Figure A-1 provides a specific illustration of the aggregate user concept. The figure shows the chain of equipment and protocol functions involved in providing communication service to a pair of application programs in a teleprocessing system. The end users in this instance are the application programs, and the user-system interfaces are between these programs and the adjacent access methods (here represented by software modules implementing the X.25 protocol). Although the end-to-end data communication system includes a number of software modules within the host computer, it may be very important for procurement purposes to define performance requirements for the digital subsystem (channel) delimited by the physical host computer interfaces. The process of allocating end-to-end performance requirements to the channel is facilitated by regarding the hosts as aggregate users of the channel, since the same performance parameters used to characterize the end-to-end service can than be used to characterize the service provided by the channel. The numerical performance values characterizing the two services will

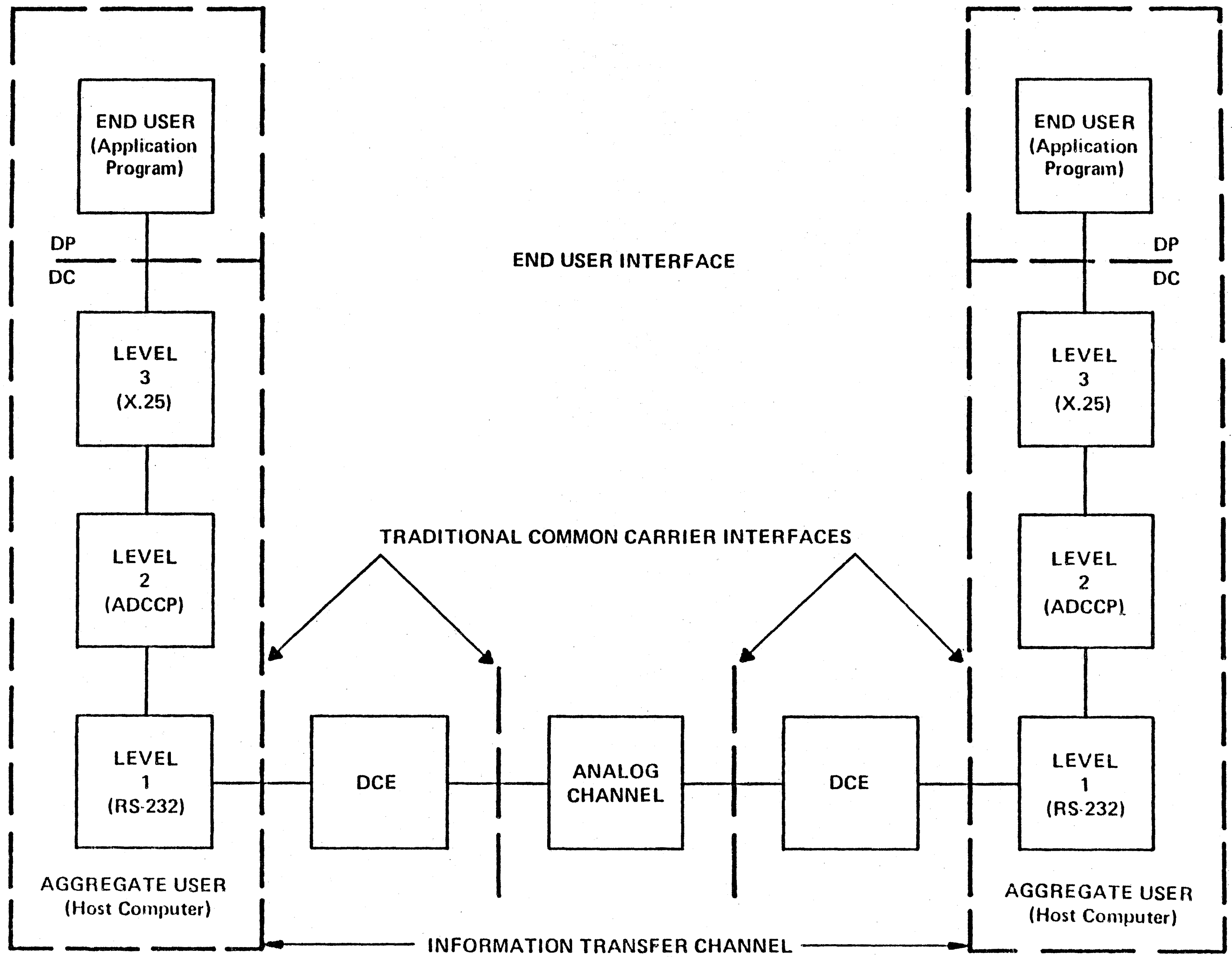


Figure A-1. Teleprocessing interfaces.

of course differ; but the difficult process of determining equivalent values for incompatibility-defined parameters is eliminated.

A-2.2 Problem Statement

Figure A-2 illustrates a typical session, or transaction, during which a pair of end users receive communication service. The transaction can be divided into three primary functional phases: access, user information transfer, and disengagement.

The access function places the system and the users in a position to begin transferring user information. It encompasses all activities normally associated with physical circuit establishment (e.g., dialing, switching, ringing) as well as any activities performed at higher protocol levels in data communication services (e.g., X.25 virtual circuit establishment).

The user information transfer function effects the actual transfer of user information between and across the user-system interfaces. It encompasses all formatting, transmission, storage, error control, and media conversion activities performed on the user information between the start of output from the source and the completion of delivery to the destination, including retransmission if required.

The disengagement function terminates the conditions that enabled user information transfer between specified users, and releases allocated system and user facilities for subsequent use. It encompasses physical circuit disconnection and resumption of line scanning in circuit switched systems, as well as higher-level protocol activities such as the termination of virtual circuits.

From the end user point of view, each individual performance of these functions can be regarded as an experiment, or "trial," which will encounter one of three general types of outcomes:

1. Successful performance. The function is performed correctly within a specified maximum performance time. For example, a voice telephone call is connected to the correct destination user within a maximum access time (e.g., 25 seconds).

2. Incorrect performance. The function is completed within the specified maximum performance time, but is performed incorrectly or unsatisfactorily. For example, a voice telephone call is connected to an unintended destination user due to a system switching error.

3. Nonperformance. The function is not completed within the specified maximum performance time. For example, a voice telephone system fails to provide dial tone to a calling user, or blocks his call with a circuit busy signal.

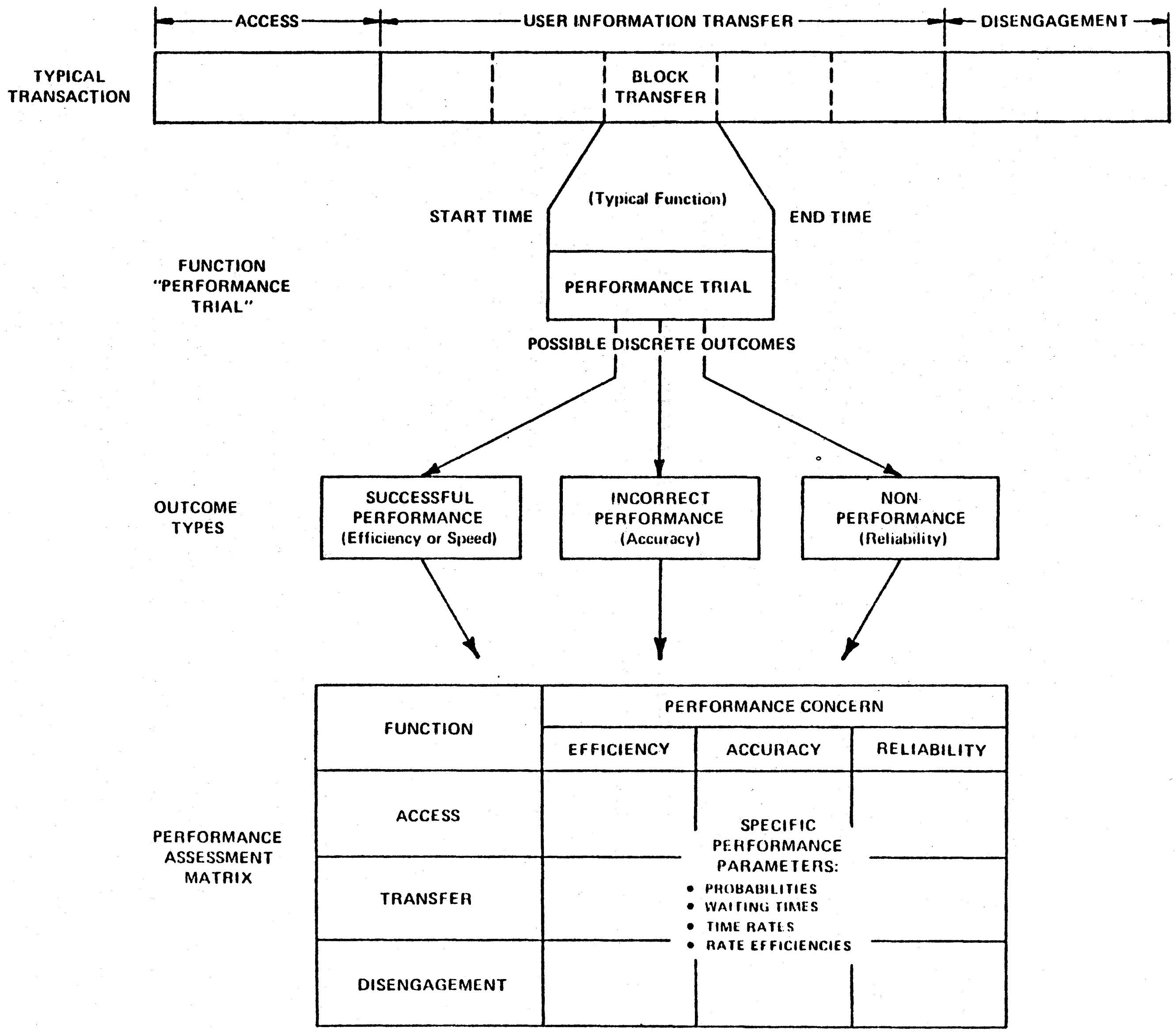


Figure A-2. Performance assessment overview.

Each of these general outcome types can be associated with a corresponding user performance concern. In the case of successful performance, the user's concern is efficiency or "speed"; in the case of incorrect or unsatisfactory performance, the user's concern is accuracy; and in the case of nonperformance, the user's concern is reliability. ITS studies indicate that these are the three most fundamental performance concerns of communications users; and that, properly interpreted, they apply to all systems and all functions performed within systems.

A complete description of performance, then, requires the separate consideration of each function of interest relative to each performance concern; and the specification of performance can be regarded as a process of "filling in the blanks" in the performing assessment matrix of Figure A-2 with specific performance parameters and values.

The above discussion provides the basis for a formal statement of the technical problem to be solved. Simply stated, the problem is to complete the expanded performance matrix of Figure A-3, as it applies to the essential communication services to be provided by DCS II. The vertical axis in this matrix distinguishes the two generic classes of communication services defined earlier: digital and analog. The functional phases of access, user information transfer, and disengagement apply, in general, to each class of service. The horizontal axis in the matrix separates the specification problem into two major parts: the development of user-oriented parameters and values, and the development of engineering-oriented (technical) parameters and values. The third axis illustrates, in general terms, the project variables associated with user class, service mode, and user interface (end-to-end or aggregate).

The brackets around the periphery of the table indicate the work to be performed during each of the three major project phases. <u>Phase A</u> will develop user-oriented parameters and values for two representative DCS II services - one digital service and one analog service. <u>Phase B</u> will translate these user-oriented parameters and values into corresponding technical parameters and values; and in so doing, will begin the process of allocating end-to-end performance requirements to component subsystems. <u>Phase C</u> will extend the criteria development process to other user classes and service modes (Figure A-4); and will allocate the resulting performance values to transmission subsystems within a defined DCS II global reference network. As noted earlier, the major emphasis of the study will be on measures which impact transmission subsystem design, i.e., the parameters and values in the <u>transfer</u> row in Figure A-4.

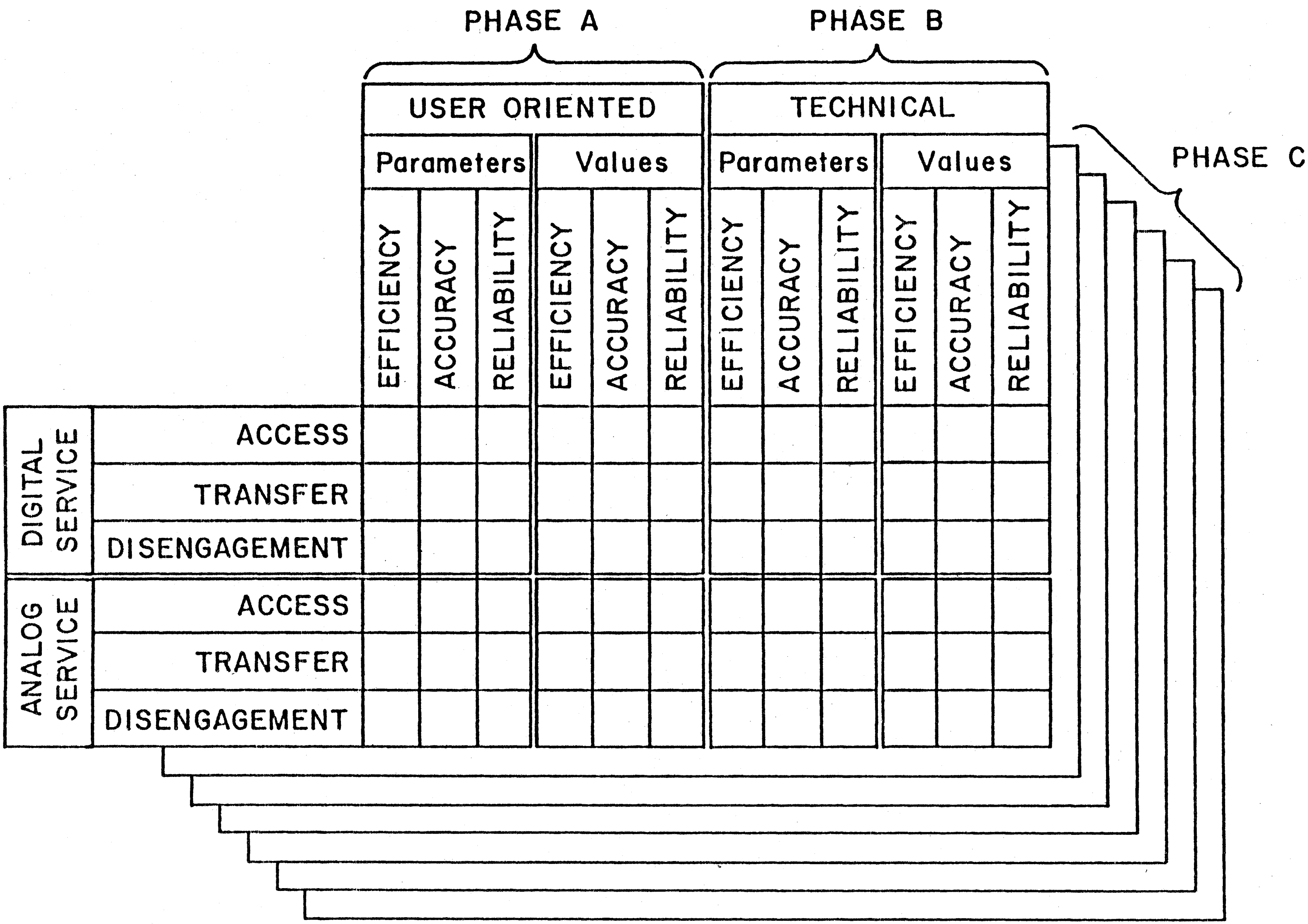


Figure A-3. Project objectives and phases.

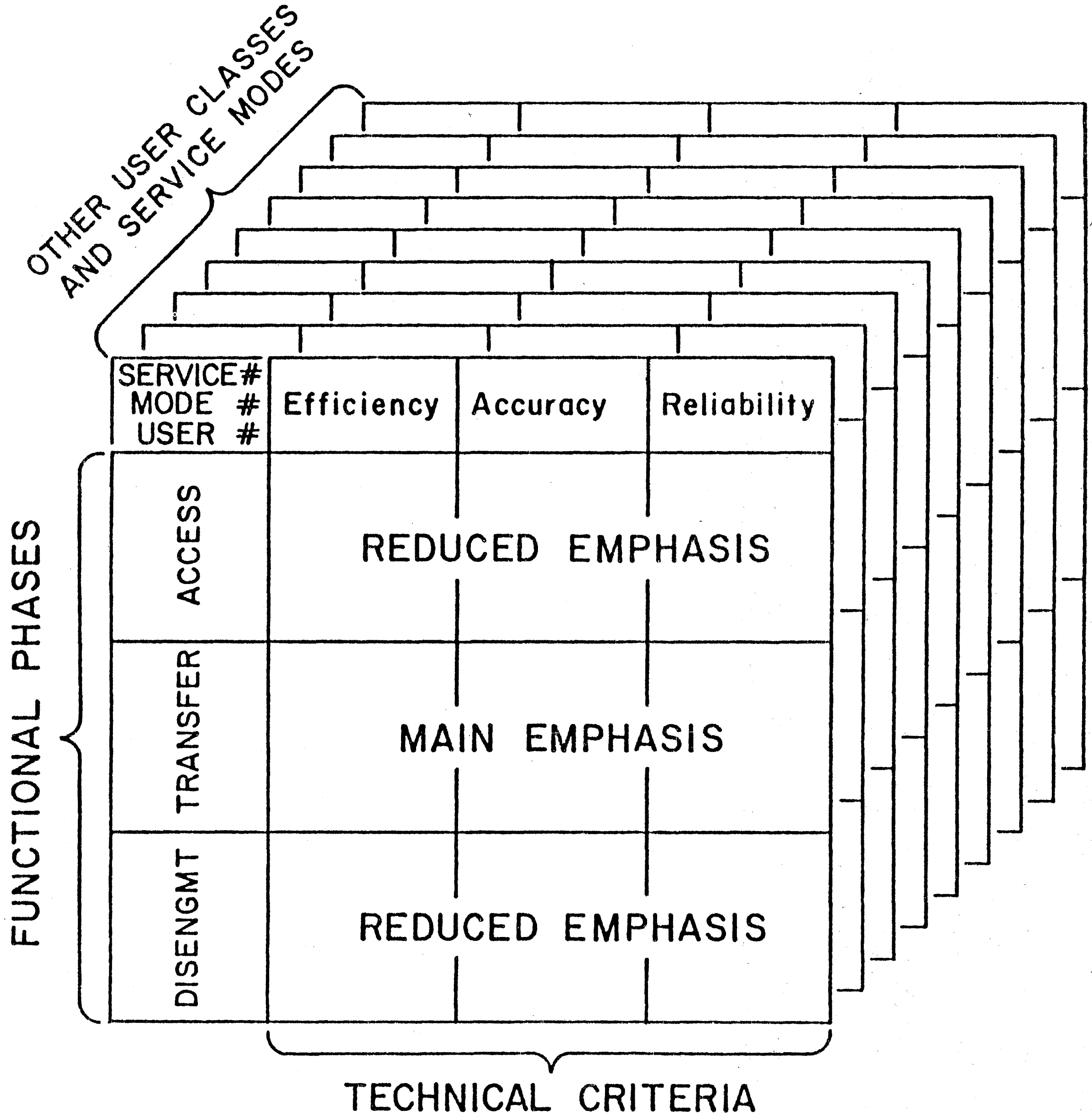


Figure A-4. Ultimate objective of Phase C.

A-3. References

Babbitt, A.E. (1977), New communications and technology in WWMCCS, Signal Magazine, August, 84-86.

DCA (1978), Defense Communications System Plan FY 1980-1990, September.

DCEC (1979), The Defense Communication System (DCS) FY 82 and Beyond, DCEC TR 20-78, March 29.

Gamauf, K.J., and W.J. Hartman (1977), Objective measurement of voice channel intelligibility, Report No. FAA-RD-77-153, October.

GAO (1977), Better management of defense communications would reduce costs, Report to the Congress by the Comptroller General of the United States, LCD 77-106, December 14.

ITS (1978), Comments on a draft work statement entitled "Determination of digital transmission technical criteria," submitted to DCEC Transmission Engineering Division, September 15.

Seitz, N.B., and P.M. McManamon (1978), Digital communication performance parameters for proposed Federal Standard 1033, NTIA Report 78-4, Volume 1, May.

APPENDIX B: A NEW FEDERAL STANDARD FOR DATA COMMUNICATION PERFORMANCE ASSESSMENT

Neal B. Seitz and Dennis Bodson*

There is a growing need within the Federal Government for a uniform means of specifying the performance of data communication systems from the point of view of the digital services delivered to the end users. This paper outlines the content and intended application of a new Federal Standard developed by the National Telecommunications and Information Administration and the National Communication System sponsored Federal Telecommunications Standards Committee to meet that need. The new standard (interim Federal Standard 1033) is unique in providing a method of describing system performance which is independent of network-internal properties such as topology and control protocol. This feature makes the standard useful both as a framework for evolving user requirements, and as a "common denominator" for comparing alternative systems or services.

B-1. Introduction

The past decade has produced two fundamental changes in the data communications industry: an enormous increase in the demand for high-quality data communication services, as a result of the growth of distributed computing; and a rapid proliferation of new sources of supply for these services, as a result of the FCC's various pro-competitive regulatory decisions. In 1970, the total market for data communication services and equipment was $600 million. Today's market is over ten times as large - $8.7 billion - and market growth in excess of 20% per year is expected through the mid-1980's. Ten years ago, a user with extensive data communication needs had essentially two choices - design, construct, and operate his own network, and maintain the necessary staff expertise in-house; or select among a limited number of available common carrier services, many ill-suited to the emerging applications. Today's user has a much wider range of options, including both enhanced service offerings of the traditional carriers and competitive offerings of the new "specialized" and "value added" carriers.

These industry changes have substantially increased the complexity - and the importance - of the communications management function. The communications manager essentially operates as a broker, or middle-man, between a user requiring communication service and supplier who provides such services. Typically, the

*Neal B. Seitz is with the Institute for Telecommunication Sciences, National Telecommunications and Information Administration, U.S. Department of Commerce, Boulder, CO 80303. Dennis Bodson is with the Office of the Manager, National Communications System, Washington, DC 20305.

user knows his application well, but has little technical knowledge of communications; conversely, the supplier knows communications well, but knows relatively little about particular user applications. The communications manager's task is to bridge the gap between these two parties, to optimally match offered systems and services with end user needs.

The consequences of inadequate or inefficient communications management can be substantial. In the absence of an effective, independent communications manager, the matching task falls to either the user or the supplier. Since the typical user has relatively little communications expertise, his "designs" tend to be inefficient, costly, brute-force approaches - e.g., dedicated lines. Suppliers also tend to over-design, both because they are uncertain about the real user need and because they receive more revenue from more elaborate services. An independent National Research Council study has estimated that the total Federal data communications bill could be reduced at least 20% through the use of more efficient methods of matching offered services with end user needs - a yearly savings in excess of $400 million by the mid-1980's (NRC, 1977). The existence of a substantial cost savings potential in the procurement of defense communications is demonstrated in a recent General Accounting Office Report (GAO, 1977).

An obvious requirement for effective communications management is a "common language" for relating the performance needs of a particular user with the performance provided by a particular system or service. A set of performance descriptors (or parameters) fulfilling this requirement would have the following general characteristics:

1. User Orientation. Selected parameters would describe the perfomance of services delivered to the end user, rather than the performance of equipment or facilities used to provide such services. The parameters would describe performance in terms of events directly observable to end users. The parameters would be chosen on the basis of performance concerns expressed by end users, rather than on the basis of engineering design considerations.

2. Universal Applicability. Selected parameters would not be restricted, in definition or application, to particular data communication systems or classes of systems. The parameter definitions would be independent of network-unique characteristics such as topology and control protocol.

3. Simplicity. Selected parameters would be simple enough to be readily understood by non-technical users; and would be defined such that users could obtain accurate measured values, within reasonable measurement intervals, using simple, inexpensive test equipment such as counters and timers. Wherever possible, parameters would be measurable during normal operational use of a service, without the need for special test scenarios.

4. Completeness. The set of selected parameters would encompass, and provide some relevant information on, all performance factors of significance to data communication users. The parameters would be "well-behaved" in the sense that they would reliably reflect actual performance over the full range of possible parameter values.

Proposed Federal Standard 1033 was developed by the National Telecommunications and Information Administration (NTIA) and the National Communications System sponsored Federal Telecommunications Standards Committee (FTSC) to provide exactly such a set of performance parameters. The FTSC voted to promulgate this proposed standard as an interim Federal Standard on March 8, 1979. The standard is currently being implemented on a trial basis within the Federal Government; and is being reviewed by a task group of the American National Standards Institute (ANSI Task Group X3S35) for later proposal as a joint Government/industry standard. It is anticipated that when these coordination steps and any necessary revisions are completed, the standard will become mandatory for use by all Federal agencies in the specification of end-to-end data communication services.

B-2. Key Technical Problems

Development of the standard required the solution of three key technical problems. The first of these was the problem of system dependence. A survey of candidate parameters revealed that the great majority were defined such that they could only be applied to systems with particular topology or protocol features. This is undesirable because it prevents use of the parameters in comparing systems that provide the same ultimate service by means of different detailed designs.

A familiar example of a system dependent parameter is the telephone parameter Dial Tone Delay - the average waiting time between the user action of going "off-hook" and the system response of providing dial tone. This parameter is useful in assessing the performance of a dial telephone network; but it has no clear meaning in the case of (for example) an ARPA-type packet switched system. Users wishing to compare circuit and packet switched services on a common performance basis would not be able to do so in terms of Dial Tone Delay. A second example of a system dependent parameter in common use is the telephone switching parameter Time to Receipt of Audible Ringing.

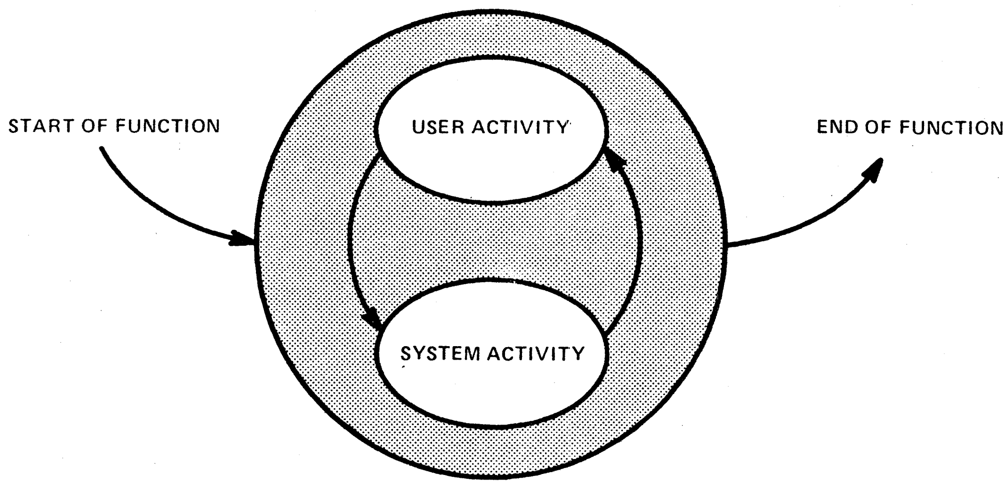
The second key problem was the problem of detailed parameter definition. In most cases, traditional narrative definitions of performance parameters are not precise enough to ensure their uniform application to comparable service offerings. The result, of course, is a potential for inefficiency and error in the process of matching service offerings with end user needs.

As an example of the parameter definition problem, consider the familiar accuracy parameter Bit Error Probability. A typical narrative defines this parameter as "bits in error per bits transmitted", but makes no mention of whether (or how) bits lost in transmission should be counted. This ambiguity can have a substantial influence on measured parameter values; for example, Bell System measurements indicate that the probability of character loss exceeds the probability of character error by more than an order of magnitude on some low-speed data links (AT&T, 1971).

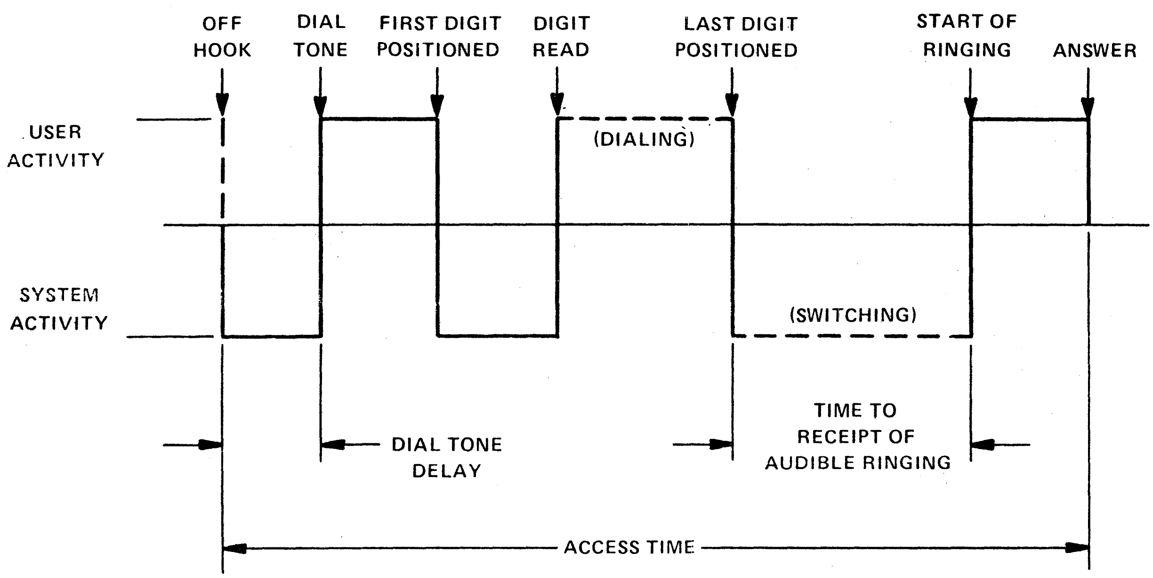
The third key problem was the problem of user dependence. In most cases, the communication process involves a sequence of interactions between the users and the system; and overall communication performance depends, then, on user performance as well as system performance. There is an obvious problem in employing user dependent parameters in specifying required system performance: the carrier or other supplier normally has no control over user performance, and hence cannot ensure that user dependent parameter values will be met. Nevertheless, many of the parameters which best describe communication performance are user dependent (e.g., "throughput").

As a simple illustration of the user dependence problem, consider the position of a user who wishes to place a voice call over the public switched network (Figure B-1). As he initiates the call, his major concern is with how soon conversation can begin, i.e., the total delay between his off-hook action and the called party's answer. The performance parameter Access Time describes exactly this delay; but its values depend not only on the system's speed in signalling and switching, but on the user's speed in dialing and answering.

The telephone companies have traditionally avoided this problem by focusing on parameters which describe unilateral system performance, e.g., Dial Tone Delay and Time to Receipt of Audible Ringing. Unfortunately, such parameters



a. Alternating Activities within a User Dependent Function



b. Voice Telephone Access Illustration

Figure B-1. User dependence in the measurement of access performance.

have two major disadvantages from the user point of view: (1) they are system dependent, as noted above; and (2) they do not reflect differences in the "functional burden" placed on the user by otherwise equivalent services. As an example, neither parameter would reflect, in terms of better performance values, the significant advantage of abbreviated dialing over conventional rotary dialing. Recent telephone company studies have recognized this limitation, and stress the need to express the influence of user performance on overall end-to-end performance (e.g., see Duffy and Mercer, 1978). Nevertheless, a precise quantitative framework for expressing this influence has not been proposed heretofore.

B-3. Federal Standard 1033 Approach

Figure B-2 summarizes the overall approach used in developing performance parameters for the standard. The parameter development process consisted of four major steps:

1. Model Development. Existing and proposed data communication services were surveyed and certain universal performance characteristics shared by all were identified. These characteristics were consolidated in a simple, user-oriented model which provided a system-independent basis for the performance parameter definitions.

2. Function Definition. Five primary communication functions were selected and defined in terms of model reference events. These functions (access, bit transfer, block transfer, message transfer, and disengagement) provided a specific focus for the parameter development effort.

3. Failure Analysis. Each primary function was analyzed to determine the possible outcomes an individual "trial performance" might encounter. Possible outcomes were grouped into three general outcome categories: successful performance, incorrect performance, and nonperformance. These categories correspond to the three general performance concerns (or "criteria") most frequently expressed by end users: efficiency, accuracy, and reliability.

4. Parameter Selection. Each primary function was considered relative to each performance criterion in matrix fashion; and one or more specific parameters were selected to represent performance relative to each function/criterion pair. Parameters were selected on the basis of expressed user interest, and consisted of probabilities, waiting

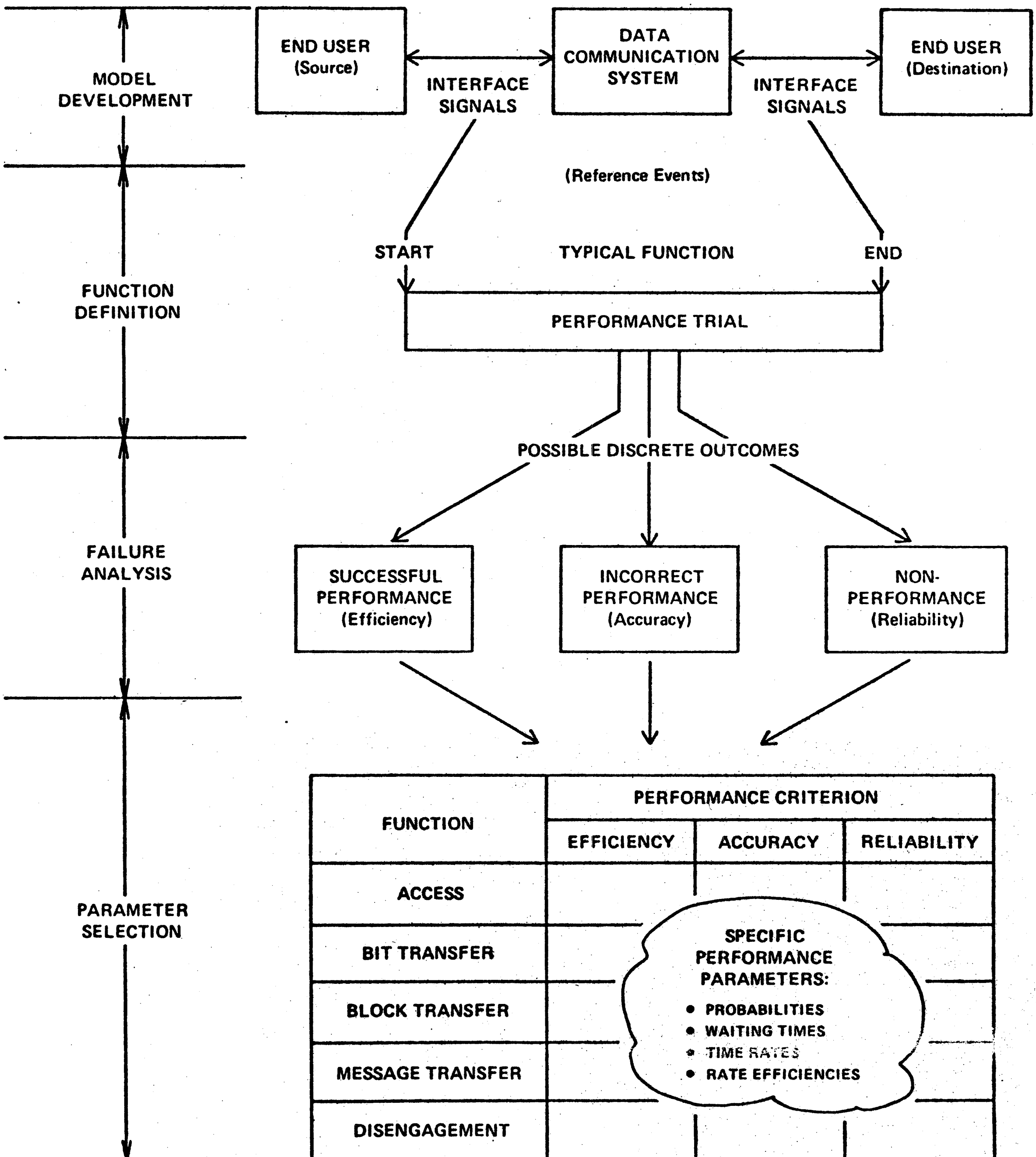


Figure B-2. Parameter development overview.

times, time rates, and rate efficiencies. The matrix approach ensured that no significant aspect of communication performance would be overlooked in the parameter selection process.

The performance model used in defining the FED STD 1033 parameters differs from those used in earlier standards and specifications in two major respects. The first is the definition of user/system interfaces. The model defines the end user of a data communication system or service as a human terminal operator, unattended device medium (e.g., punched cards), or computer application program. Thus, the end-to-end data communication system extends to the operator or medium side of the data terminal, or to the applications program side of the host computer telecommunications access method[1]. The end-to-end system includes higher-level communication control protocols such as ANSI's ADCCP (FED STD 1003) and CCITT's X.25.

This viewpoint is essential in a user-oriented standard, since modern terminals and protocols perform functions (such as error control, flow control, and virtual circuit establishment) which have a profound effect on end-to-end performance. One modern data communication network whose end user interfaces are defined in this way is IBM's Systems Network Architecture (McFadyen, 1976). The International Standards Organization is developing a reference model for distributed computer networks which follows a similar interface definition approach (ISO, 1978).

The second major difference between the FED STD 1033 model and earlier performance models is in the selection of parameter defining events. In any description of data communication performance, certain information transfers or device state changes are identified as events to be counted, timed or compared in calculating performance parameter values. As noted earlier, most existing standards and specifications identify such events by reference to particular system-dependent interface signals (e.g., off-hook). The FED STD 1033 model departs from this approach by defining the performance parameters in terms of more general, system independent reference events. Each FED STD 1033 reference event is a "generic event" which subsumes many system-specific interface signals having a common performance significance; and each is defined in such a way that

---

[1]A "telecommunications access method" is a program in a teleprocessing application computer which serves as the first point of contact for local application programs requiring telecommunication service. One example of such a program is IBM's Virtual Telecommunications Access Method (Albrecht and Ryder, 1976).

it necessarily occurs, at some point, in any end-to-end data communication transaction.

System-specific interface signals are mapped into FED STD 1033 reference events on the basis of the user interface involved, the type of information transferred (e.g., user information or overhead), and the nature of the state change the transfer produces. One example of a system-independent reference event defined in FED STD 1033 is the start of block transfer. If a block of information is to be moved between end users, it must at some point pass from the physical possession and control of the source user to that of the system. The identity of the information unit called a block, and the physical method of transfer, will vary from system to system; but the event start of block transfer can be identified in any system. A more rigorous state-machine presentation of the FED STD 1033 model is provided in Seitz and McManamon (1978).

Any description of performance ultimately refers to some particular function. The five primary communication functions considered in FED STD 1033 were defined in terms of particular model events as follows.

> The access function begins on issuance of an Access Request signal at the originating user interface, and ends (successfully) on the first subsequent transfer of a user information bit or block from a source user to the system. It encompasses all activities traditionally associated with physical circuit establishment (e.g., dialing, switching, ringing, modem handshaking) as well as any activities performed at higher protocol levels (e.g., X.25). Making the end of access coincident with the start of user information transfer reflects the user view that no data communication service has actually been provided until user information begins to flow.

> The bit, block, and message transfer functions describe the flow of information between end users at three distinct levels of detail. Each function begins on the start of output of the associated information unit from the source user, and ends (successfully) on completion of delivery of that unit to the intended destination. Each function encompasses all formatting, transmission, storage, error control, and media conversion activities performed between start of output and completion of delivery, including retransmission if required. All three functions must be considered in a comprehensive performance specification, as discussed below.

The disengagement function begins on issuance of a Disengagement Request signal at either user interface, and ends (successfully) on return of a corresponding Disengagement Confirmation signal.

As noted earlier, the terms Access Request, Disengagement Request, and Disengagement Confirmation are general descriptions of purpose (generic events) rather than specific names of interface signals. They denote, respectively, any event whose purpose is to initiate, terminate, or confirm termination of an entity's participation in an information transfer transaction.

The bit transfer, block transfer, and message transfer functions each serve a distinct purpose in the description of user information transfer performance. The bit transfer function fulfills the need for a "common denominator" to enable performance comparison between services having different block lengths: performance parameters can always be compared at the bit level. The block transfer function focuses attention on the information unit that is most relevant to the user in his internal operations - the user information block[2]. The message transfer function provides a formal basis for defining the so-called "secondary" parameters, which describe the long-term availability of a data communication service. The "message" information unit is basically a sample size, determined on the basis of measurement confidence limits as described in Crow and Miles (1976).

In conducting the failure analysis, we regarded each individual performance of a primary function as an experiment, or trial, in the statistical sense; and defined a set of possible outcomes of each "performance trial" in a pie diagram of outcome possibilities called a sample space. Figure B-3 depicts six possible outcomes of an individual performance of a general primary function g (representing any one of the five). The individual outcomes are distinguished on the basis of the expected ending event (and any unexpected intermediate events) as follows:

1. Successful Performance (g_s). The expected ending event occurs, and is correct in both location (user interface) and content (delivered information).

---

[2]As used in the standard, the term "block" denotes a contiguous group of user information bits delimited at the source user/system interface for transfer to a destination user as a unit. Thus, for instance, a block may be a single ASCII character, a card image, a computer word, or the information field of a frame, depending on the equipment and protocol characteristics at the user/system interface.

2. Content Error (g_e). The expected ending event occurs at the correct location, but is incorrect in content.

3. Location Error (g_m). The expected ending event, or a required intermediate event, occurs at an incorrect location.

4. System Nonperformance (g_ℓ). The expected ending event does not occur within the maximum performance period, as a result of either issuance of a blocking signal or excessive delay on the part of the telecommunication system.

5. User Nonperformance (g_f). The expected ending event does not occur within the maximum performance period, as a result of either issuance of a blocking signal or excessive delay on the part of a user.

6. Extra Event (g_x). A nonblocking event, not included within the expected event sequence, occurs.

These outcomes are grouped into three categories in the sample space of Figure B-3, in accordance with the three user performance concerns (or criteria) noted earlier: efficiency, accuracy, and reliability.

As a specific example, consider the application of these outcome categories to the function of block transfer. Successful performance is the case where the source block is delivered to the intended destination within a specified maximum block transfer time, and is correct in content. Content error is the case where the delivered block contains one or more bit errors, additions, or deletions. Location error is the case where the source block is delivered to an incorrect (unintended) destination. System nonperformance is the case where the source block is lost in transmission (e.g., as a result of an acknowledgement error). User nonperformance is the case where the source block is not delivered as a result of a failure of the destination user to perform some required action (e.g., allocate necessary buffer space). Extra event is the case where the system delivers a block not output by any source (e.g., a duplicate block). Figure B-4 summarizes the sample spaces for the five primary functions by indicating the general outcome categories included in each.

The final step in the FED STD 1033 parameter development process was to select and define specific parameters to represent system performance relative to each function/criterion pair. Figure B-5 illustrates how this was accomplished in the case of the access function. Access performance was described in terms of three specific parameters, one associated with each of the three general performance criteria noted earlier.

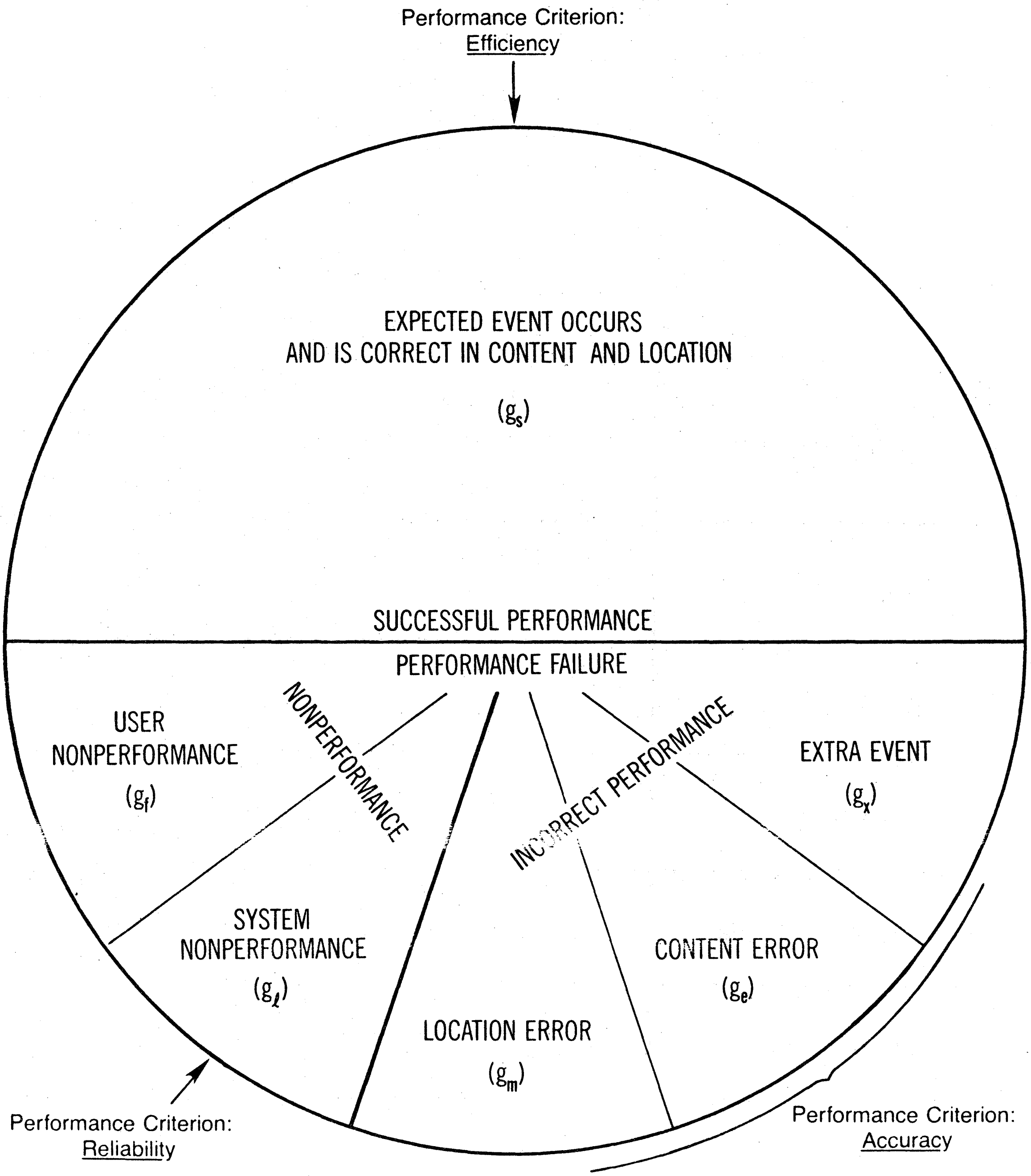
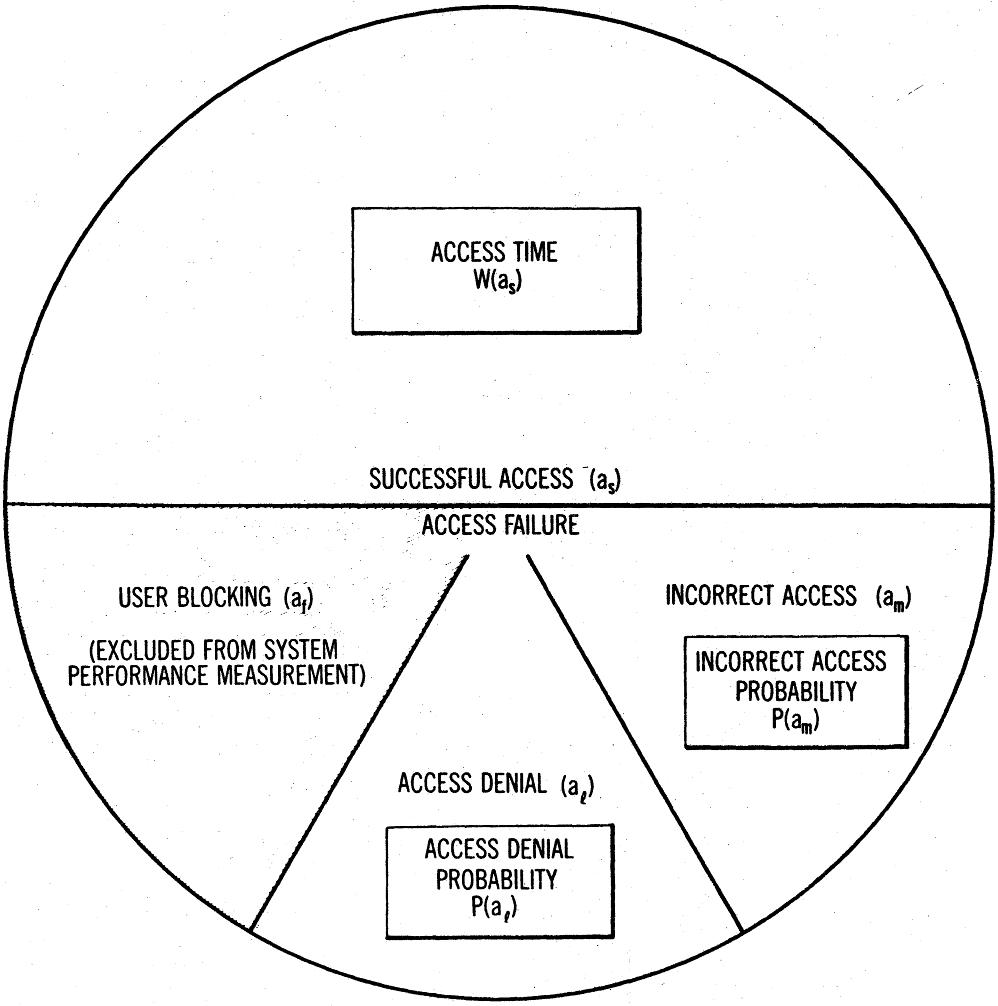


Figure B-3. Sample space for a typical primary function.

PRIMARY FUNCTIONS	OUTCOMES INCLUDED IN SAMPLE SPACE					
	g_s	g_e	g_m	g_ℓ	g_f	g_x
ACCESS	√		√	√	√	
BIT TRANSFER	√	√	√	√	√	√
BLOCK TRANSFER	√	√	√	√	√	√
MESSAGE TRANSFER	√	√	√	√	√	√
DISENGAGEMENT	√			√	√	

Figure B-4. Primary function outcomes.



ACCESS PARAMETERS

1. Access Time $= W(a_s) = \frac{1}{A_s} \sum_{a_s=1}^{A_s} w(a_s)$
2. Incorrect Access Probability $= P(a_m) = A_m/A'$
3. Access Denial Probability $= P(a_\ell) = A_\ell/A'$

DEFINITIONS

A_s = Total number of Successful Access outcomes counted during an access parameter measurement.

A_m = Total number of Incorrect Access outcomes counted during an access parameter measurement.

A_ℓ = Total number of Access Denials counted during an access parameter measurement.

A' = Total number of access attempts counted during an access parameter measurement: $A_s+A_m+A_\ell$.

$w(a_s)$ = Value of access time measured on a particular successful access attempt.

Figure B-5. Access parameter definitions.

1. Access Time - Average value of elapsed time between the start of an access attempt and Successful Access. Elapsed time values are calculated only on access attempts that result in Successful Access.

2. Incorrect Access Probability - Ratio of total access attempts that result in Incorrect Access (i.e., connection to an unintended destination) to total access attempts counted during a parameter measurement.

3. Access Denial Probability - Ratio of total access attempts that result in Access Denial (i.e., system blocking) to total access attempts counted during a parameter measurement.

A maximum access time equal to three times the specified value of the parameter Access Time was defined as a "timeout constant" for performance measurement purposes; access attempts not completed within this maximum time are counted as access failures. The access failure probabilities are calculated on the basis of a reduced measurement sample which excludes access attempts which fail due to user blocking (e.g., the called party is busy or does not answer).

A key aspect of the FED STD 1033 parameter definitions is their expression in mathematical form (Figure B-5). This approach eliminates the ambiguity associated with traditional narrative definitions, and ensures that the parameters will be applied in a consistent way in all situations.

The same general approach used in the access case was followed in selecting and defining performance parameters for the user information transfer and disengagement functions. A separate probability parameter was defined to express the likelihood of each possible failure outcome; and the successful performance outcomes were expressed in terms of waiting times and (in the case of bit and block transfer) time rates. Bit and block transfer rate efficiencies were also defined, to express system performance from the standpoint of resource utilization.

B-4. Secondary and Ancillary Parameters

Although the primary parameters described above provide a relatively detailed description of data communication performance, they fall short of completeness in two respects:

1. They do not provide the kind of macroscopic, long-term performance view users traditionally associate with the concept of availability.

2. They are user dependent, and thus cannot be applied directly in situations where it is necessary to describe unilateral system performance.

A small set of additional "secondary" and "ancillary" performance parameters were included in the standard to meet these needs.

Figure B-6 illustrates the approach used in defining the secondary (availability) parameters. The sequence of transmissions between a specified pair of users is divided into a series of consecutive performance measurement periods, each corresponding to the number of transmitted bits in a "message" as defined above. On completion of each successive message transfer function, values for each of five "supported" primary performance parameters are calculated. The calculated values are compared with corresponding outage thresholds to define the outcome of that trial performance of the message transfer function as either Operational Service state or Outage state. A service is defined to have been in the Operational Service state (during the preceding performance measurement period) whenever the measured values for all supported parameters are better than their associated outage thresholds; a service is defined to have been in the Outage state whenever the measured values for one or more supported parameters are worse than their associated thresholds.

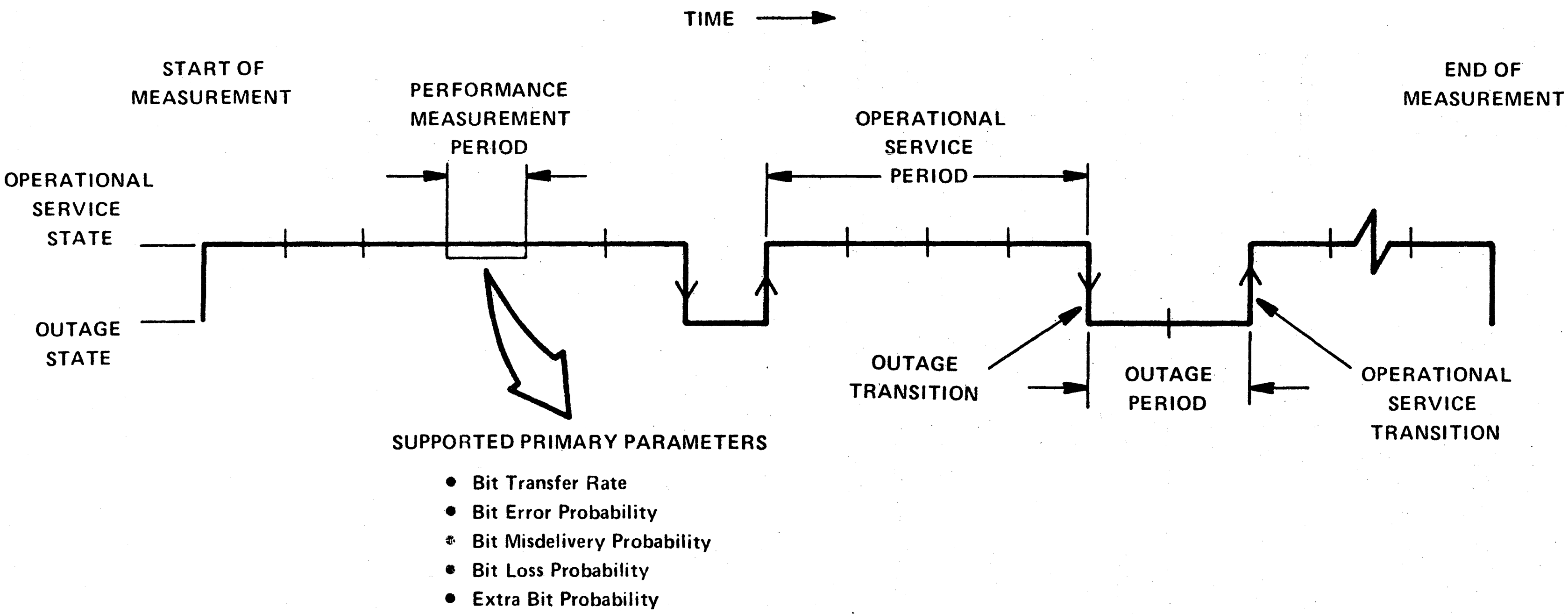
Five primary user information transfer parameters are defined as supported performance parameters: the four bit transfer failure probabilities (Bit Error Probability, Bit Misdelivery Probability, Bit Loss Probability, and Extra Bit Probability) and Bit Transfer Rate. Outage thresholds for the supported performance parameters are defined as a function of the corresponding nominal values (specified for the service) as follows:

1. The outage threshold for Bit Transfer Rate is defined as one-third (1/3) of the nominal Bit Transfer Rate.

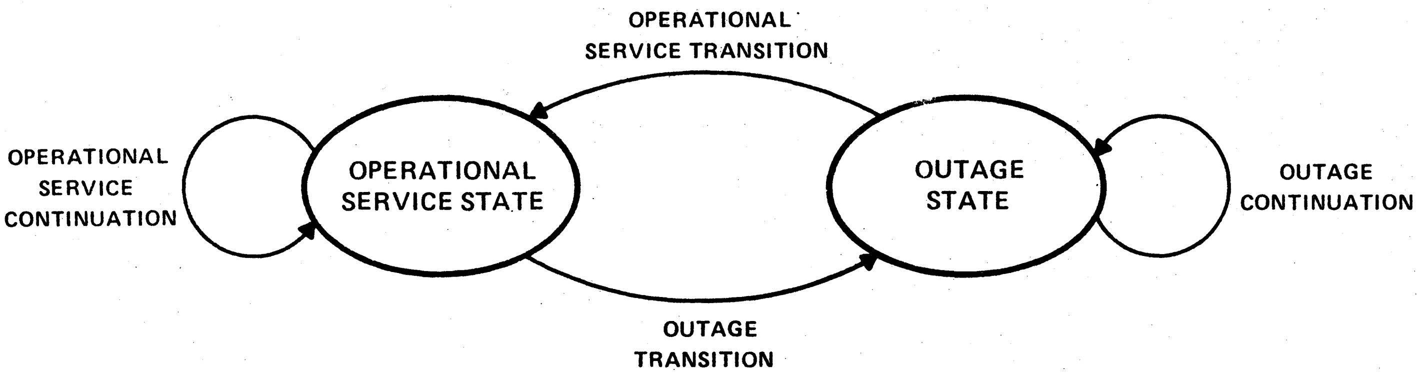
2. The outage thresholds for the four bit transfer failure probabilities are defined as a function of the corresponding nominal probability values by expressing the nominal value as a power of ten (for example, 10^{-6}) and then dividing the exponent by two (producing, for example, a threshold value of 10^{-3}).

Three secondary performance parameters are defined on the basis of the two-state (Markov) outage model shown in Figure B-6b:

Service Time Between Outages - Average value of elapsed time between entering and next leaving the Operational Service state.



a. Secondary State Sequence



b. Secondary State Diagram

Figure B-6. Secondary performance parameters.

Outage Duration - Average value of elapsed time between entering and next leaving the Outage state.

Outage Probability - Ratio of total message transfer attempts resulting in the Outage state to total message transfer attempts included in a measurement sample.

Service Time Between Outages and Outage Duration are equivalent to the traditional availability parameters Mean Time Between Failures (MTBF) and Mean Time to Repair (MTTR). Outage Probability is essentially a sampled measure of unavailability. These parameters are termed "secondary" to emphasize the fact that they are defined on the basis of measured primary parameter values, rather than on the basis of direct observations of interface events.

The "ancillary" parameters provide a quantitative means of expressing the influence of user delays on the primary parameter values. As noted earlier, the primary functions are defined in such a way that they normally involve a sequence of interactions between the users and the system (Figure B-1); and their overall performance time thus depends on both user and system delays. In essence, the FED STD 1033 approach divides this overall performance time into user and system fractions, and defines the average user fraction as an ancillary parameter which modifies or "qualifies" the associated primary parameters.

Four ancillary parameters are defined in the standard, each expressing the average user fraction of the total performance time for an associated primary function[3]. As an example, the ancillary parameter User Access Time Fraction is defined as the ratio of average user access time to average total Access Time measured over a sample of successful access attempts.

Each defined ancillary parameter can be used directly as a correction factor, to calculate "user-independent" values for the associated primary efficiency parameters. If $W(g_s)$ is the specified performance time for a primary function and $p(g)$ is the associated ancillary parameter, the user-independent time for the function is

$$[1p(g)] \cdot W(g_s).$$

The factor $[1-p(g)]$ is the average system performance time fraction - the complement of $p(g)$. Similarly, given any specified rate or rate efficiency parameter

[3]No ancillary parameter is defined for the bit transfer function since its values can be inferred from the corresponding block transfer parameter.

value, $R(g_s)$ or $Q(g_s)$, the corresponding user-independent value can be calculated as

$$\frac{R(g_s)}{[1-p(g)]} \quad \text{or} \quad \frac{Q(g_s)}{[1-p(g)]} \; .$$

In each case, the user-independent values express the performance that would be provided by the system if all user delays were zero; i.e., if all user activities were performed in zero time. As an example, assume the Access Time value for the service of Figure B-1 is specified as 25 seconds, with an associated User Access Time Fraction of 0.6. Then the user-independent Access Time value - the average total system delay during access - is (0.4)(25)=10 seconds.

The ancillary parameters also provide a basis for identifying the entity "responsible" for timeout failures; e.g., whether an access attempt not completed within the maximum access time should be attributed to Access Denial or User Blocking (Figure B-5). This decision is made by comparing the ancillary parameter value characterizing the particular trial in question with the corresponding <u>average</u> ancillary parameter value. If the user fraction for the particular trial exceeds the corresponding average, the failure is attributed to the user; otherwise, the failure is attributed to the system. This application of the ancillary parameters is described more fully in Seitz and McManamon (1978).

B-5. Problem Solutions - Summary

We noted earlier that the development of FED STD 1033 required the solution of three key technical problems. The technical approach adopted in the standard provides a solution to each of these problems, as summarized below.

1. <u>System Dependence</u>. The standard solves this problem through the expedient of the user-oriented performance model. The model reduces all user/system interactions to a small set of general <u>reference events</u> which can be identified in any system; and the performance parameter definitions are then based on these system-independent events.

2. <u>Detailed Parameter Definition</u>. The standard solves this problem by using sample spaces and mathematical equations as the major parameter definition tools. <u>Sample spaces</u> encourage the analyst to consider, and carefully define, all relevant outcomes of a performance trial. <u>Equation definitions</u> eliminate the ambiguity often associated with purely narrative definitions.

3. User Dependence. The standard solves this problem through the use of the ancillary performance parameters. These parameters provide a basis for "factoring out" user influence on the waiting time, time rate, and rate efficiency parameters; and a means of determining whether the user or the system is "responsible" for timeout failures.

Figure B-7 summarizes the performance parameters ultimately selected for inclusion in FED STD 1033. A total of 26 parameters were selected, including 19 primary parameters, 3 secondary parameters, and 4 ancillary parameters.

B-6. Intended Application

In conclusion we present a feasible, if somewhat optimistic, view of the ultimate application of FED STD 1033; and a summary of steps currently being taken to approach that goal. An ideally effective application of the standard might be described in terms of the following general scenario:

1. Federal user organizations would understand and accept the need to describe their communication requirements in a functional, system-independent manner. Users would specify performance requirements in FED STD 1033 terms, without reference to particular communication facilities or services. Individual parameter values would be determined on the basis of their impact on the user process being served; as an example, the Bit Error Probability requirement for a digital air traffic control system would be determined by considering the impact of bit errors on air traffic control effectiveness.

2. Federal communications managers would have the authority and the resources needed to select the best means of meeting user requirements. Where appropriate, they would aggregate independent user requirements to be met by a single common user system. They would develop procurement specifications by allocating end-to-end performance objectives to purchasable subsystems in terms of the FED STD 1033 parameters.

3. Industry suppliers of data communication systems and services would be willing to specify their performance in FED STD 1033 terms, and would be appropriately compensated for their effort in doing so. Available facilities and services would be catalogued, along with major Federal applications, in a central communications data base which would facilitate the design and procurement process.

FUNCTION	PERFORMANCE CRITERION			
	EFFICIENCY	ACCURACY	RELIABILITY	
ACCESS	• ACCESS TIME	• INCORRECT ACCESS PROBABILITY	• ACCESS DENIAL PROBABILITY	• USER ACCESS TIME FRACTION
BIT TRANSFER	• BIT TRANSFER TIME	• BIT ERROR PROBABILITY • BIT MISDELIVERY PROBABILITY • EXTRA BIT PROBABILITY	• BIT LOSS PROBABILITY	• USER BLOCK TRANSFER TIME FRACTION
BLOCK TRANSFER	• BLOCK TRANSFER TIME	• BLOCK ERROR PROBABILITY • BLOCK MISDELIVERY PROBABILITY • EXTRA BLOCK PROBABILITY	• BLOCK LOSS PROBABILITY	
MESSAGE TRANSFER	• BIT TRANSFER RATE • BLOCK TRANSFER RATE • BIT RATE EFFICIENCY • BLOCK RATE EFFICIENCY	• OUTAGE PROBABILITY		• USER MESSAGE TRANSFER TIME FRACTION
DISENGAGEMENT	• DISENGAGEMENT TIME	• DISENGAGEMENT DENIAL PROBABILITY		• USER DISENGAGEMENT TIME FRACTION
SERVICE CONTINUATION	• SERVICE TIME BETWEEN OUTAGES			
SERVICE RESTORAL	• OUTAGE DURATION			

Legend

Primary Parameters

Secondary Parameters

Ancillary Parameters

Figure B-7. Summary of selected performance parameters.

4. User, manager, and service supplier would monitor delivered performance using the FED STD 1033 parameters as needed to (a) verify system compliance with key requirements; (b) support effective network control; and (c) monitor actual usage patterns. Performance measurements would be facilitated by standard procedures and software tailored to the FED STD 1033 parameters[4].

The successful application of FED STD 1033 would provide a substantial cost savings to the Federal Government in the specification, procurement, and operation of Federal data communication systems. Requirements specification would be simplified by the standardization of performance descriptors, and by the existence of a data base of similar previous applications. Procurement would be improved by the standard's "common denominator" property, and by the facility and service data bases. Operations would be enhanced by more effective network control. The $400 million figure cited earlier provides an indication of the total Federal cost savings potential; clearly, many similar benefits would accrue to non-Federal users of the standard.

The major organizational tools needed to achieve the benefits cited above exist within the Federal government at the present time. Interim Federal Standard 1033 is one of a series of standards being developed under the auspices of the Federal Telecommunications Standards Committee (FTSC). This interagency committee, operating under the sponsorship of the Office of the Manager of the U.S. government's National Communication System, is authorized by GSA and the Executive Office of the President to develop telecommunication standards for Federal-wide use (Bodson, 1978); and GSA has the authority to require that Federal Agencies comply with these standards. Federal Property Management Regulations published by GSA's Automated Data and Telecommunications Service (ADTS) already define the general factors to be considered in conducting data communication requirements studies, and point out that GSA assistance in the conduct of such studies is available (GSA, 1978). Thus, GSA ADTS has both the authority and the responsibilities of an independent communications manager with respect to most Federal data communication procurements. Additional technical resources are provided by the Federal Information Processing Standards Program of the National Bureau of Standards.

---

[4]Such tools are being developed for a new proposed Federal Standard FED STD 1043).

Existence of the necessary organizational machinery does not, of course, ensure that the transition from existing practice to full realization of the benefits cited above will be trouble-free. While the standard has undergone considerable critical scrutiny, its operational use has not been extensively tested; and initial applications will undoubtedly reveal some areas which should be clarified or revised[5]. Even if no such areas exist, a substantial effort will still be required to educate Federal users and communications managers in its use. Finally, industry suppliers will have to be convinced of its benefits and provided with appropriate implementation incentives.

In recognition of these needs, the FTSC has elected to pursue a gradual approach in implementing the new standard. This approach involves three specific steps:

1. Approval of proposed Federal Standard 1033 on an interim (optional) basis.

2. Development and publication of a FED STD 1033 user guide, plus additional application examples.

3. Joint development, with ANSI Task Group X3S35, of an American National Standard based on FED STD 1033.

Approval of the standard on an optional basis will allow Federal agencies maximum flexibility in selecting initial applications, and will facilitate any revisions suggested by such trials. The user guide and examples will familiarize users with the standard, and will demonstrate its practical application. The joint Federal/ANSI development effort will bring additional industry expertise to bear on the problem, and will promote industry understanding and acceptance of the user-oriented approach. As noted earlier, it is anticipated that these steps will ultimately lead to mandatory application of the standard in all Federal procurements.

A major determinant of the success of any standard is the breadth of participation in its development. In order to ensure the broadest possible consensus for Federal Standard 1033, interested readers are encouraged to suggest improvements and, if possible, to contribute directly to the ANSI TG 5 effort. Copies of the standard and its supporting reports can be obtained by writing the authors at the addresses noted above.

---

[5]The standard has been applied to three representative systems from a specification viewpoint (Kimmett and Seitz, 1978); and a subset of the defined parameters have been measured over the ARPA network (Payne, 1978).

B-7. References

Albrecht, H.R., and K.D. Ryder (1976), The virtual telecommunications access method: A systems network architecture perspective, IBM Systems Journal, Volume 15, No. 1.

AT&T (1971), Bell System Technical Reference Pub. 41007, 1969-70 switched telecommunications network connection survey, April.

Bodson, Dennis (1978), The National Communications System, IEEE Communications Society Magazine, March.

Crow, E.L., and M.J. Miles (1976), A low-cost, accurate statistical method to measure bit error rates, Proceedings of the International Conference on Computer Communications, Toronto, Canada, August.

Duffy, F.P., and R.A. Mercer (1978), A study of network performance and customer behavior during direct-distance-dialing call attempts in the USA, Bell System Technical Journal, Vol. 57, January.

GAO (1977), Better management of defense communications would reduce costs, Report to the Congress by the Comptroller General of the United States, LCD-77-106, December 14.

ISO (1978), Reference model of open systems architecture, version 3 International Standards Organization, TC97/SC16 N117, November.

Kimmett, F.G., and N.B. Seitz (1978), Digital communication performance parameters for proposed Federal Standard 1033, Volume II, Application examples, May.

NRC (1977), National Research Council Committee on Telecommunications, summary of Office of Telecommunications Study Panel meeting, Boulder, CO, April 11-13.

McFadyen, J.H. (1976), Systems network architecture: an overview, IBM Systems Journal, Vol. 15, No. 1.

Payne, J.A. (1978), ARPANET host-to-host access and disengagement measurements, NTIA Report 78-3, May.

Seitz, N.B., and P.M. McManamon (1978), Performance parameters for proposed Federal Standard 1033, Volume I, standard parameters, NTIA Report 78-4, May.

APPENDIX C: LINEAR PREDICTIVE CODING

Linear predictive coding (LPC) has long been used in communication theory. More recently, it has found applications in speech analysis and synthesis, speaker identification, and word recognition, to name just a few new areas.

LPC models the vocal tract as an all-pole digital filter and estimates the filter parameters (predictor coefficients) using the time domain speech waveform itself, rather than the waveform's short-term frequency spectrum. This makes LPC a relatively efficient method for encoding speech compared to frequency domain techniques.

The vocal tract is assumed to be modeled as a discrete, time-varying filter with parameters changing slowly enough so that they can be considered fixed over a specified time interval. Hence, the vocal tract can be approximated by a series of stationary shapes. Atal and Hanauer (1971) have shown that this all pole model can account for the glottal volume flow and radiation of sound from the mouth in addition to vocal tract sounds.

The transfer function, H(z), used to describe the digital model over each analysis frame is given by

$$H(z) = \frac{1}{1 - \sum_{i=1}^{P} a_i z^{-i}} \quad . \tag{C-1}$$

for a model with P poles.

The time sequence S_n corresponding to the output of the recursive filter can be written as

$$S_n = \sum_{i=1}^{P} a_i \, S_{n-i} + \delta_n \qquad n = 0, 1, \ldots \tag{C-2}$$

where (a_i) are the predictor coefficients that completely describe the characteristics of the filter and (δ_n) is the driving function or input to the filter.

While there have been several formulations for the estimation of the linear prediction coefficients, two least squares methods have become prominent, the autocorrelation method and the co-variance method. The autocorrelation method will be used in this study. The autocorrelation method can be considered as estimating the filter coefficients by approximating the spectrum of the speech waveform by an all-pole model.

The portion of the signal to be analyzed is first multiplied by a finite window of length N, changing the signal to

$$S_n = \begin{cases} \text{Windowed speech sampled,} & o \le n \le N-1 \\ \\ o, & n < o \text{ and } n \ge N \end{cases} \tag{C-3}$$

Using this windowed signal, the prediction error sequence is defined as

$$e_n = S_n - \hat{S}_n = S_n - \sum_{i=1}^{P} a_i S_{n-i} \qquad n = 0, 1, \ldots \tag{C-4}$$

and the total squared error is then

$$E_T = \sum_n e_n^{\,2} = S_o^{\,2} + \sum_{n=1}^{N-1+P} (S_n - \sum_{i=1}^{P} a_i S_{n-i})^2. \tag{C-5}$$

The predictor coefficients are selected so as to minimize the total squared error. This is accomplished by setting the partial derivative of the total squared error with respect to each predictor coefficient equal to zero. The system of equations that results is

$$\sum_{k=1}^{P} r_{|i-k|} a_k = r_i \qquad i = 1, 2, \ldots, P \tag{C-6}$$

where

$$r_i = \frac{1}{R_o} \sum_{n=o}^{N-1-|i|} S_n S_{n+|i|} \qquad i = 0, 1, 2, \ldots, P \tag{C-7}$$

are the normalized short-term autocorrelation values of the speech signal and

$$R_o = \sum_{n=o}^{N-1} S_n^{\,2} \tag{C-8}$$

is the normalization factor for these values. The normalized total squared error can be defined by making use of equations (C-5) and (C-6), yielding

$$E = E_{min} = 1 - \sum_{i=1}^{P} a_i r_i \quad . \tag{C-9}$$

The predictor coefficients are obtained by inverting a positive definite Toeplitz matrix with elements

$$r_{|i-k|} \qquad i, k = 1, 2, \ldots, P \tag{C-10}$$

Further insight can be gained by looking at the frequency domain approximation to the above system. Taking the z-transform of equation (C-4), one obtains

$$E(z) = S(z)\,[H(z)]^{-1} \tag{C-11}$$

where H(z) is defined in equation (C-1) and E(z) and S(z) are the z-transforms of e_n and S_n respectively. Rearranging, equation (C-11) can be written as

$$S(z) = E(z)\,H(z). \tag{C-12}$$

Minimizing the total squared prediction error is equivalent to approximating the error sequence, (e_n), by

$$\hat{e}_n = \begin{cases} A\ , & n = 0 \\ 0\ , & n \neq 0 \end{cases} \tag{C-13}$$

in the least squares sense. This implies that E(z) is being approximated by the function A, a constant, and S(z) is being approximated by a spectrum corresponding to an all-pole transfer function, i.e.,

$$\hat{E}(z) = A \tag{C-14}$$

$$\hat{S}(z) = \hat{E}(z)\,H(z) = \frac{A}{1 - \sum_{i=1}^{P} a_i z^{-i}} \tag{C-15}$$

The value of A is determined by the application of energy conservation between $\hat{e}_n$ and e_n. Using equations (C-9) and (C-13), one obtains

$$A^2 = E_{min} = E = 1 - \sum_{i-1}^{P} a_i \, r_i \qquad \text{(C-16)}$$

thereby showing that A^2 is equal to the minimum total squared error of the system. Figure C-1 shows the relationship between S_n, E, and S_n', E'.

This approach of estimating filter coefficients so as to minimize the energy of the output of the inverse of a system driven by its impulse response is sometimes called deconvolution or inverse filtering. Considerable work has been done in this area of linear prediction of speech in the past few years and many good references are available that give more detailed discussions of this subject. Some particularly good ones are Markel and Gray (1976 and 1973); Makhoul (1975 and 1973); Makhoul and Wolf (1972); and Boll (1973).

C-1. Time Domain Interpretations

This total squared error or linear prediction residual energy can be considered to be the output of an inverse filter $H(z)^{-1}$, where

$$[H(z)]^{-1} = 1 + \sum_{i=1}^{P} a_i \, z^{-1} \quad . \qquad \text{(C-17)}$$

$[H(z)]^{-1}$ is the filter that minimizes the residual energy, E. H(z) corresponds to a smoothed spectral estimate of the data sequence (S_n) up to a scale factor representing the gain. The relationship between S_n, E, and $[H(z)]^{-1}$ is illustrated in Figure C-2. $[H'(z)]^{-1}$ is the filter that minimizes the residual energy E'. H'(z) corresponds to a smoothed spectral estimate of the data sequence S_n' up to a scale factor representing the gain. The relationship between S_n', E', and $[H'(z)]^{-1}$ is also illustrated in Figure C-2.

If (S_n) is passed through a different inverse filter, $[H'(z)]^{-1}$, of the form

$$[H'(z)]^{-1} = 1 + \sum_{i=1}^{P} a_i' \, z^{-1} \qquad \text{(C-18)}$$

which minimizes the residual energy (E') for some other data sequence (S_n'), the residual energy D, must be greater than or equal to the minimum residual energy E, i.e., $D \geq E$, with the equality holding if and only if H(z) = H'(z). Assuming the data sequence (S_n) is obtained from an analysis frame of speech from the input

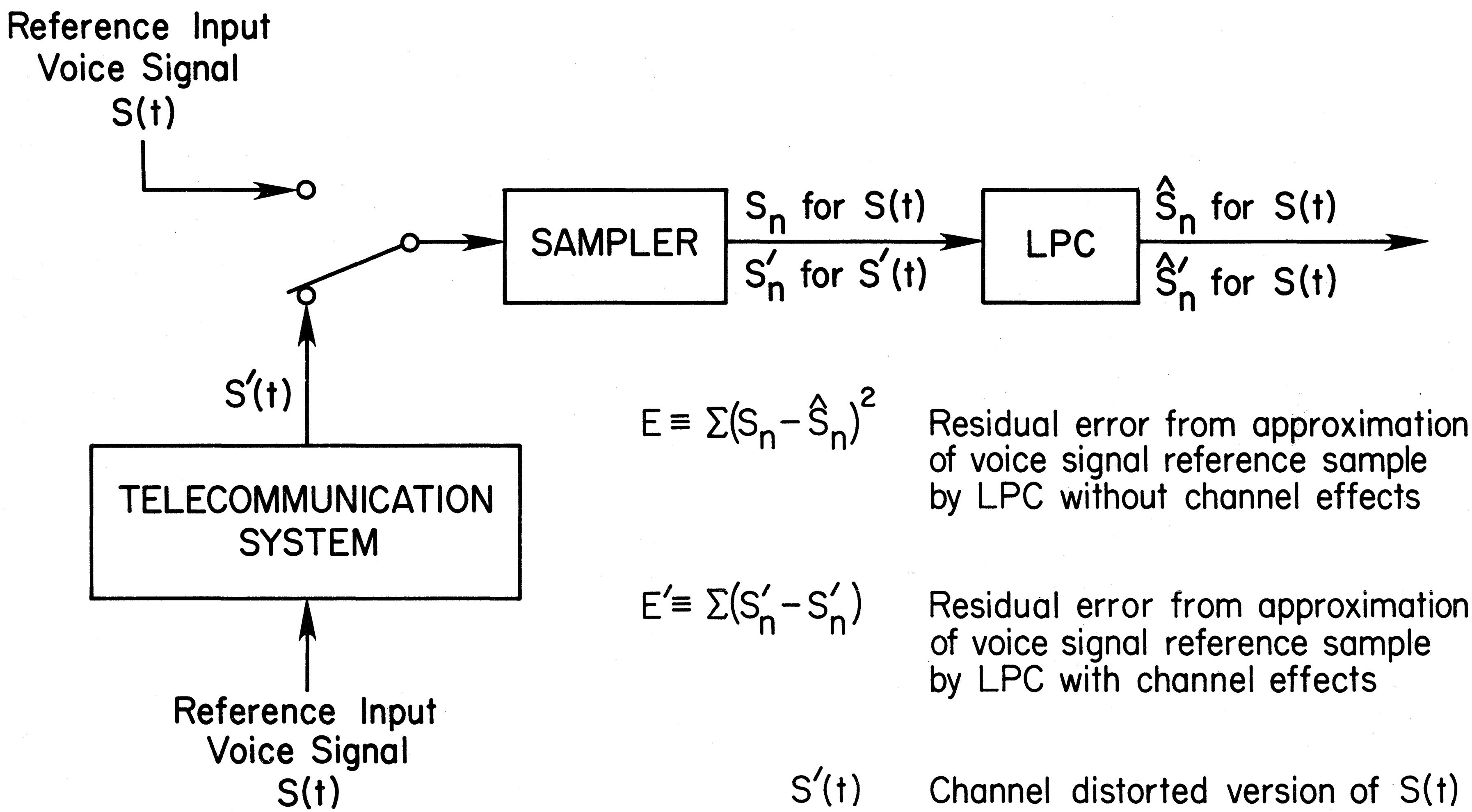


Figure C-1. Summary of total squared error measures E and E'.

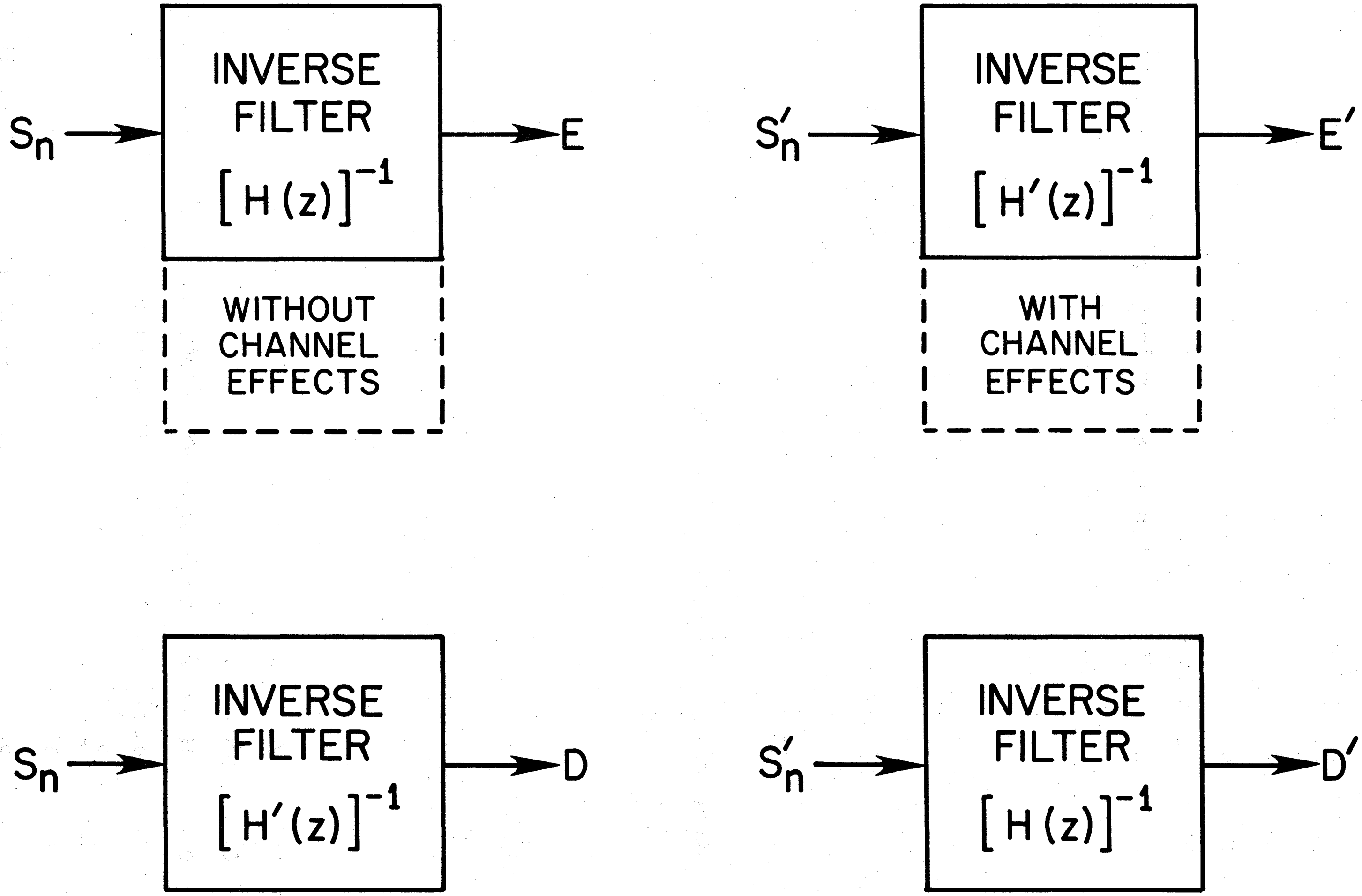


Figure C-2. Relationship between D, D', E, and E'.

and (S_n') is the corresponding analysis frame from the output, the difference between D and E is a measure of the distance between the two speech segments. (Unless otherwise identified, unprimed variables represent data from the input.)

The dual of the above situation is also true. If (S_n') is sent through $[H(z)]^{-1}$, the output will be D', while E' is the output of $[H'(z)]^{-1}$ when (S_n') is sent through it. Again, $D' \geq E'$ with equality if and only if $H'(z) = H(z)$. As before, the difference between D' and E' can be considered to be a distance measure between the two speech segments.

E, E', D, and D' can all be written as a combination of the autocorrelation terms of (S_n) and (S_n') and the corresponding linear prediction coefficients (a_i) and (a_i'). Let

$$\underline{A}^T = (1, -a_i, -a_2, \dots -a_p) \qquad \text{(C-19)}$$

be the transpose of the linear prediction coefficient vector $\underline{A}$, and

$$\underline{\underline{R}} = r_{|i-k|} \qquad i, k = 0, 1, 2, \dots, P \qquad \text{(C-20)}$$

the normalized autocorrelation matrix. The four error terms can then be written as

$$E = \underline{A}^T \underline{\underline{R}} \underline{A} \qquad \text{(C-21)}$$

$$E' = \underline{A}'^T \underline{\underline{R}}' \underline{A}' \qquad \text{(C-22)}$$

$$D = \underline{A}'^T \underline{\underline{R}} \underline{A}' \qquad \text{(C-23)}$$

$$D' = \underline{A}^T \underline{\underline{R}}' \underline{A} \qquad \text{(C-24)}$$

where the primes signify variables from the output. The derivation of this can be found in Markel and Gray (1973).

E and E' are calculated for each analysis frame using the Levinson Algorithm, but D and D' are calculated when the output is compared to the input because of the structure of $\underline{\underline{R}}$ and $\underline{\underline{R}}'$, the autocorrelation matrices. These calculations can be simplified by calculating the autocorrelation terms of $\underline{A}$ and $\underline{A}'$, the linear prediction coefficient vectors. Using the symmetry of the autocorrelation terms of $\underline{A}$, and $\underline{A}'$, D and D', can be written as

$$D = \sum_{i=0}^{P} g_i \, r_i \qquad \text{(C-25)}$$

$$D' = \sum_{i=0}^{P} g_i \, r_i' \tag{C-26}$$

where

$$q_i = 2 \cdot \sum_{k=0}^{P-|i|} \alpha_k \, \alpha_{k+i} \qquad (i = 1, 2, \ldots, P \tag{C-27}$$

$$g_o = \sum_{k=0}^{P} \alpha_k^2 \tag{C-28}$$

$$g_i' = 2 \cdot \sum_{k=0}^{P-|i|} \alpha_k' \, \alpha_{k+1}' \qquad (i = 1, 2, \ldots, \tag{C-29}$$

$$g_o' = \sum_{k=0}^{P} \alpha_k'^2 \quad . \tag{C-30}$$

(α_i) and (α_i') are the P+1 terms in the vectors $\underline{A}$ and $\underline{A'}$ respectively.

C-2. Frequency Domain Interpretations

Each of the four error terms can be interpreted in the frequency domain. Using Parseval's Theorem, the total squared error, E, can be written as

$$E = \sum_n e_n 2 = \frac{T}{2\pi} \int_{-\pi/T}^{\pi/T} |E(\omega)|^2 \, d\omega \quad , \tag{C-31}$$

where $E(\omega)$ is obtained by substituting $z = e^{i\omega t}$ into $E(z)$. From (C-11), the minimum linear prediction error was found to be

$$E(z) = S(z) \, [H(z)]^{-1} \tag{C-32}$$

while the least squares estimate can be written as

$$\hat{E}(z) = \hat{S}(z) \, [H(z)]^{-1} \tag{C-33}$$

substituting $z = e^{i\omega t}$ into (C-32) and (C-33) one obtains

$$E(\omega) = S(\omega) \, [H(\omega)]^{-1} \tag{C-34}$$

$$\hat{E}(\omega) = \hat{S}(\omega) \, [H(\omega)]^{-1} \quad . \tag{C-35}$$

Rearranging (C-35) and substituting in $\hat{E}(z) = A$,

$$[H(\omega)]^{-1} = \frac{A}{\hat{S}(\omega)} \quad . \tag{C-36}$$

Inserting (C-36) into (C-34),

$$E(\omega) = A \frac{S(\omega)}{\hat{S}(\omega)} \quad . \tag{C-37}$$

substituting (C-37) into (C-31)

$$E = \frac{T\,A^2}{2\pi} \int_{-\pi/T}^{\pi/T} \frac{|S(\omega)|^2}{|\hat{S}(\omega)|^2}\, d\omega \tag{C-38}$$

But $|S(\omega)|^2$ and $|\hat{S}(\omega)|^2$ are just the corresponding power spectra, $P(\omega)$ and $\hat{P}(\omega)$, of the speech signal and its least squares linear prediction estimate. Therefore,

$$E = \frac{TA^2}{2\pi} \int_{-\pi/T}^{\pi/T} \frac{P(\omega)}{\hat{P}(\omega)}\, d\omega \quad . \tag{C-39}$$

Similarly E' can be shown to be

$$E' = \frac{TA'^2}{2\pi} \int_{-\pi/T}^{\pi/T} \frac{P'(\omega)}{\hat{P}'(\omega)}\, d\omega \quad . \tag{C-40}$$

D and D' can also be obtained by the same method and written as

$$D = \frac{TA'^2}{2\pi} \int_{-\pi/T}^{\pi/T} \frac{P(\omega)}{\hat{P}'(\omega)}\, d\omega \tag{C-41}$$

$$D' = \frac{TA^2}{2\pi} \int_{-\pi/T}^{\pi/T} \frac{P'(\omega)}{\hat{P}\,(\omega)}\, d\omega \quad . \tag{C-42}$$

The distance measures D and D' are not all that pleasing when defined in the frequency domain. They compare the ratios between a true speech power spectrum and an estimated power spectrum. A much more desirable measure would compare the ratios between the estimated power spectra of the input speech and the output version of it. This can be done by taking the ratio of D to E and D' to E'. The ratio of each of these pairs of residual errors, D/E and D'/E', then defines new distance measures which are more appropriate. In both cases, the ratios are greater than or equal to one, with equality if and only if $H(z) = H'(z)$.

The ratios D/E and D'/E' are sometimes called likelihood ratios because under certain circumstances, they have been shown to be true likelihood ratios by Itakura (1975). As mentioned before, the frequency domain interpretation of the likelihood ratios gives a good justification for using them as distance measures. In the time domain

$$\frac{D}{E} = \frac{\sum_n D_n^2}{\sum_n e_n^2} \quad , \tag{C-43}$$

where

$$\sum_n D_n^2 = \sum_n \left(S_n - \sum_{i=1}^{P} a_i' \, S_{n-i} \right)^2 \tag{C-44}$$

$$\sum_n e_n^2 = \sum_n \left(S_n - \sum_{i=1}^{P} a_i \, S_{n-i} \right)^2 \tag{C-45}$$

Gray and Markel (1976) have shown that D/E can be written in the frequency domain as

$$\frac{D}{E} = \frac{T}{2\pi} \int_{-\pi/T}^{\pi/T} \frac{|H(\omega)|^2}{|H'(\omega)|^2} \, d\omega \quad , \tag{C-46}$$

where the substitution $z = e^{j\omega T}$ is made in the filters $H(z)$ and $H'(z)$. Inverting (C-36) and its dual, one obtains

$$H(\omega) = \frac{\hat{S}(\omega)}{A} \; , \quad \text{and} \tag{C-47}$$

$$H'(\omega) = \frac{\hat{S}'(\omega)}{A'} \quad . \tag{C-48}$$

Substituting (C-47) and (C-48) into (C-46)

$$\frac{D}{E} = \frac{T}{2\pi} \int_{-\pi/T}^{\pi/T} \frac{|S(\omega)|^2}{A^2} \quad \frac{A'^2}{|\hat{S}'(\omega)|^2} \, d\omega \, , \tag{C-49}$$

and

$$\frac{D}{E} = \frac{TA'^2}{2\pi A^2} \int_{-\pi/T}^{\pi/T} \frac{|S(\omega)|^2}{|\hat{S}'(\omega)|^2} \, d\omega \quad . \tag{C-50}$$

Once again, the magnitude squared of the signal's spectrum is just its power spectrum, therefore

$$\frac{D}{E} = \frac{TA'^2}{2\pi A^2} \int_{-\pi/T}^{\pi/T} \frac{\hat{P}(\omega)}{\hat{P}'(\omega)} \, d\omega \tag{C-51}$$

Similarly

$$\frac{D'}{E'} = \frac{TA^2}{2\pi A'^2} \int_{=\pi/T}^{\pi/T} \frac{\hat{P}'(\omega)}{\hat{P}(\omega)} \, d\omega \tag{C-52}$$

As can be seen from (C-51) and (C-52), D/E and D'/E' compute the ratio between the estimated power spectra of the reference and channel-distorted speech. While D and D' compare the true power spectra of the reference speech to estimates of the power spectra of the channel-distorted speech.

The LPC analysis requires the calculation of the autocorrelation functions R_i, i = 1, 2, . . . P, for each frame. Solving the sets of equations (C-6) by recursion results in the following procedure.

$$E_o = R_o$$

$$k_i = - [R_i + \sum_{j=1}^{i-1} a_i^{(i-1)} R_{i-j}]/E_{i-1} \tag{C-53}$$

$$a_i^{(i)} = k_i$$

$$a_j^{(i)} = a_j^{(i-1)} + k_i \, a_{i-j}^{(i-1)} \, , \; 1 \le j \le i - 1$$

$$E_i = (1 - k_i^2) \, E_{i-1}.$$

The error E_i is calculated at each step and $0 \le E_i \le E_{i-1}$. The k_i are known as either reflection coefficients or Parcor coefficients.

One additional set of parameters of interest are the area coefficients, defined by

$$A_i = \frac{1 + k_i}{1 - k_i} \quad 1 \le i \le P \tag{C-54}$$

C-1. References

Atal, B.S., and S.L. Hanauer (1971), Speech analysis and synthesis by linear prediction of the speech wave, J. of the Acoust. Society of America, 50, 637-655, April.

Boll, S.F. (1973), A priori digital speech analysis, Ph.D thesis, University of Utah, March.

Gray, A.H., Jr., and J.D. Markel (1976), Distance measures for speech processing, IEEE Trans. on Acoustics, Speech, and Signal Processing, ASSP-24, 380-391, October.

Itakura, F. (1975), Minimum prediction residual principle applied to speech recognition, IEEE Trans. on Acoustics, Speech, and Signal Processing, ASSP-23, 67-72, February.

Makhoul, J. (1973), Spectral analysis of speech by linear prediction, IEEE Trans. on Audio and Electroacoustics, AU-21, 140-148, June.

Makhoul, J. (1975), Linear prediction: a tutorial review, Proceedings of the IEEE, 63, 561-580, April.

Makhoul, J.I., and J.J. Wolf (1972), Linear prediction and the spectral analysis of speech, (Bolt, Beranek, and Newman, Inc. Report No. 2304, Cambridge).

Markel, J.D., and A.H. Gray, Jr. (1973), On autocorrelation equations as applied to speech analysis, IEEE Trans. on Audio and Electroacoustics, AU-21, 69-79, April.

Markel, J.D., and A.H. Gray, Jr. (1976), Linear prediction of speech, (New York: Springer-Verlag).

FORM **NTIA-29**
(4-80)

U.S. DEPARTMENT OF COMMERCE
NAT'L. TELECOMMUNICATIONS AND INFORMATION ADMINISTRATION

BIBLIOGRAPHIC DATA SHEET

	1. PUBLICATION NO.	2. Gov't Accession No.	3. Recipient's Accession No.
	NTIA Report 81-57		

4. TITLE AND SUBTITLE	5. Publication Date
Performance Parameters for Digital and Analog Service Modes	January 1981
	6. Performing Organization Code: NTIA/ITS-4
7. AUTHOR(S): M. Nesenbergs, W. J. Hartman, and R. F. Linfield	9. Project/Task/Work Unit No.
8. PERFORMING ORGANIZATION NAME AND ADDRESS: U. S. Department of Commerce, NTIA/ITS-4, 325 Broadway, Boulder, Colorado 80303	10. Contract/Grant No.
11. Sponsoring Organization Name and Address: Defense Communications Agency, Defense Communications Engineering Center, Reston, Virginia 22090	12. Type of Report and Period Covered
	13.

14. SUPPLEMENTARY NOTES

15. ABSTRACT *(A 200-word or less factual summary of most significant information. If document includes a significant bibliography or literature survey, mention it here.)*

As with many technologies, the evolution of telecommunication systems is shaped by two driving forces - performance and cost. There is a real need to bridge the gap between 'performance' as perceived by the user in accomplishing a mission and 'performance' as perceived by the supplier to minimize costs of implementation and operations. The interrelationships between user-oriented performance parameters, engineering-oriented parameters, and cost parameters ultimately define the permissible tradeoffs.

This report covers one phase of a multiple phase project to relate the performance needs of military network users with the performance provided by a particular telecommunications service and to seek least cost systems that would offer such service.

In this first phase, the parameters describing the performance of two services, one analog and one digital, are defined and specific values assigned for the related service offerings. The interrelationship between the user-oriented parameters and values and the technical or engineering-oriented parameters and values was planned as the subject of a subsequent phase.

Key Words: analog communications; digital communications; performance parameters; performance standards; service modes; system design; user requirements

17. AVAILABILITY STATEMENT	18. Security Class. *(This report)*	20. Number of pages
X☒ UNLIMITED.	Unclassified	294
☐ FOR OFFICIAL DISTRIBUTION.	19. Security Class. *(This page)* Unclassified	21. Price:

★ U. S. GOVERNMENT PRINTING OFFICE 1980 - 778-563/212 Reg. 8