NBS TECHNICAL NOTE 882

U.S. DEPARTMENT OF COMMERCE / National Bureau of Standards

Criteria for the Performance Evaluation of Data Communications Services for Computer Networks

NATIONAL BUREAU OF STANDARDS

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

THE INSTITUTE FOR BASIC STANDARDS provides the central basis within the United States of a complete and consistent system of physical measurement; coordinates that system with measurement systems of other nations; and furnishes essential services leading to accurate and uniform physical measurements throughout the Nation's scientific community, industry, and commerce. The Institute consists of a Center for Radiation Research, an Office of Measurement Services and the following divisions:

Applied Mathematics — Electricity — Mechanics — Heat — Optical Physics — Nuclear Sciences[2] — Applied Radiation[2] — Quantum Electronics[3] — Electromagnetics[3] — Time and Frequency[3] — Laboratory Astrophysics[3] — Cryogenics[3].

THE INSTITUTE FOR MATERIALS RESEARCH conducts materials research leading to improved methods of measurement, standards, and data on the properties of well-characterized materials needed by industry, commerce, educational institutions, and Government; provides advisory and research services to other Government agencies; and develops, produces, and distributes standard reference materials. The Institute consists of the Office of Standard Reference Materials and the following divisions:

Analytical Chemistry — Polymers — Metallurgy — Inorganic Materials — Reactor Radiation — Physical Chemistry.

THE INSTITUTE FOR APPLIED TECHNOLOGY provides technical services to promote the use of available technology and to facilitate technological innovation in industry and Government; cooperates with public and private organizations leading to the development of technological standards (including mandatory safety standards), codes and methods of test; and provides technical advice and services to Government agencies upon request. The Institute consists of a Center for Building Technology and the following divisions and offices:

Engineering and Product Standards — Weights and Measures — Invention and Innovation — Product Evaluation Technology — Electronic Technology — Technical Analysis — Measurement Engineering — Structures, Materials, and Life Safety[4] — Building Environment[4] — Technical Evaluation and Application[4] — Fire Technology.

THE INSTITUTE FOR COMPUTER SCIENCES AND TECHNOLOGY conducts research and provides technical services designed to aid Government agencies in improving cost effectiveness in the conduct of their programs through the selection, acquisition, and effective utilization of automatic data processing equipment; and serves as the principal focus within the executive branch for the development of Federal standards for automatic data processing equipment, techniques, and computer languages. The Institute consists of the following divisions:

Computer Services — Systems and Software — Computer Systems Engineering — Information Technology.

THE OFFICE FOR INFORMATION PROGRAMS promotes optimum dissemination and accessibility of scientific information generated within NBS and other agencies of the Federal Government; promotes the development of the National Standard Reference Data System and a system of information analysis centers dealing with the broader aspects of the National Measurement System; provides appropriate services to ensure that the NBS staff has optimum accessibility to the scientific information of the world. The Office consists of the following organizational units:

Office of Standard Reference Data — Office of Information Activities — Office of Technical Publications — Library — Office of International Relations.

---

[1] Headquarters and Laboratories at Gaithersburg, Maryland, unless otherwise noted; mailing address Washington, D.C. 20234.
[2] Part of the Center for Radiation Research.
[3] Located at Boulder, Colorado 80302.
[4] Part of the Center for Building Technology.

Criteria for the Performance Evaluation of Data Communications Services for Computer Networks

Prepared December 1974

+. Technical note, no. 882.

Dana S. Grubb and Ira W. Cotton

Computer Systems Engineering Division
Institute for Computer Sciences and Technology
National Bureau of Standards
Washington, D.C. 20234

U.S. DEPARTMENT OF COMMERCE, Rogers C. B. Morton, *Secretary*
US NATIONAL BUREAU OF STANDARDS, Ernest Ambler, *Acting Director*

Issued September 1975

National Bureau of Standards Technical Note 882

Nat. Bur. Stand. (U.S.), Tech. Note 882, 36 pages (Sept. 1975)

CODEN: NBTNAE

U.S. GOVERNMENT PRINTING OFFICE
WASHINGTON: 1975

For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402 (Order by SD Catalog No. C13.46:882). Price 85 cents. (Add 25 percent additional for other than U.S. mailing).

CONTENTS

Criteria for the Performance Evaluation of Data Communications Services for Computer Networks

Dana S. Grubb and Ira W. Cotton

ABSTRACT

In general, when telecommunications services are used as a means of interchanging information between information processing systems, or between terminals and systems, a number of parameters determine how well that interchange is performed. This report examines the following characteristics of telecommunications services:

1. Transfer Rate
2. Availability
3. Reliability
4. Accuracy
5. Channel Establishment Time
6. Network Delay
7. Line Turnaround Delay
8. Transparency
9. Security

These terms are all defined and their significance discussed. The effects of these factors on data communication networks are illustrated.

Key words: Computer communications; computer networking; data communications; networks; performance requirements; telecommunications.

INTRODUCTION

In general, when telecommunications services are used as a means of interchanging information between information processing systems, or between terminals and systems, a number of parameters determine how well that interchange is performed. This report examines the following characteristics of telecommunications services:

1) Transfer Rate
2) Availability
3) Reliability
4) Accuracy
5) Channel Establishment Time
6) Network Delay
7) Line Turnaround Delay
8) Transparency
9) Security

These criteria are all user-oriented performance parameters. Thus, they differ from such operational measures as signal-to-noise ratio or signaling rate, which have traditionally been used to evaluate communications services. Users of data communications services are only interested in the external manifestations of network parameters in order to select the service offering the best performance for their particular needs and at a price they can afford to pay. For this purpose the criteria listed above are superior.

In order to design a system with the best overall performance, it is necessary to consider the relationships between all of these parameters. These parameters are not all independent and attempts to improve the performance of one parameter may degrade one or more of the others. Thus, an understanding of the inter-relationships is necessary for effective trade-offs to be made.

In this report the meanings of these communications service parameters are clarified and their impact and significance with regard to information processing systems are discussed. There is a particular attempt to demonstrate how these parameters, which were established for traditional communications services, may be applied to newer services, such as the packet-switched networks now coming into existence. Examples of the values of these parameters have been selected from both new and traditional communications services.

These parameters are not felt to be an exhaustive list for evaluating and comparing data communications services. However, they are felt to be among the most significant parameters, and the evaluation of different services on the basis of these parameters can provide a means for the comparison of quite different communications services. Differences in performance between services, according to these parameters, will often be reflected in price differences. A parametric analysis of alternative services should enable users to identify the true cost of different levels of performance, and, thereby, help them to select the service which is most cost-effective for their particular requirements.

1. TRANSFER RATE

1.1 Definition

The number of information bits transferred by means of the telecommunications channel divided by the time required for their transfer and acceptance.

1.2 Other Definitions

Transfer rate is defined in ANSI X3.44 - 1974 (for data communication systems using the X3.28 - 1971 character-oriented control procedure) as Transfer Rate of Information Bits (TRIB). (Ref. 1 & 2.) TRIB is defined as "The ratio of the number of Information Bits accepted by the receiving Terminal Configuration during a single Information Transfer Phase (Phase 3) to the duration of the Information Transfer Phase. TRIB is expressed in bits per second." (The capitalized terms have specific meanings that are defined in the standard.)

1.3 Significance of Transfer Rate

Signaling rate is the data rate of the telecommunications service in bits per second. However, the computer user is concerned with the number of information bits that are transferred in a unit of time. The transfer rate is reduced below that of the telecommunication service's signaling rate by such factors as the bits used for the header and the error detection code, and the retransmission of message blocks or frames containing errors. Thus, the user may specify a particular signaling rate from the common carrier, but the actual transfer rate will usually be substantially less. Increasing the signaling rate on a given telecommunications facility will usually (though not always) increase the transfer rate, but the increase in the transfer rate is often less than the increase in the signaling rate due to the increased effect of the factors cited above.

1.4 Discussion

Signaling rates are generally offered by the common carriers or by the vendors of equipment for use on common carrier facilities in increments specified in the ANSI standard X3.1 - 1969 (ref. 3) for synchronous transmission and at lower speeds for asynchronous transmission. Line quality is the concern of the common carriers. It affects the user by the degree to which it reduces the transfer rate by requiring the retransmission of information blocks or frames containing errors caused by poor line quality. Control procedures (hardware and software protocol used to transfer data and control information) are the concern of the information processing system staff. The choice of the control procedure to be used involves software and hardware considerations, but also affects the transfer rate by the amount of overhead bits that must be allocated for control and for error detection and retransmission purposes. As discussed in section 4.3, the undetected error rate may be made arbitrarily small by increasing the number of bits used for error detection, but at the cost of increasing the overhead and reducing the transfer rate.

Propagation delay and line turnaround (if applicable) are factors that affect the transfer rate and, with line quality, affect the choice of block or frame length to be used for an optimal transfer rate. (Block or frame length will be limited by the nature of the usage of the system and by the available storage for buffering.)

Where sophisticated control procedures are employed for routing or guaranteeing the integrity of data, measuring the transfer rate may result in some differing opinions. To the communications engineer, control procedures which are implemented as headers and trailers on message blocks or frames are part of the data. However, to the applications-oriented computer specialist, they are overhead. In a sophisticated computer network (such as, ARPANET) with many levels of control protocol, all implemented in software, even the computer specialists may not agree on the specific criteria to be used to measure the effective transfer rate. It is essential that a consistent interpretation be followed when comparisons are made between services.

1.5 Typical Transfer Rate Requirements of Information Processing Systems

Typical transfer rate experience with interactive terminals on various computer networks are shown in the table below, with all speeds shown in bits per second (ref. 4). The user´s rates are limited by the user´s think time and typing speed and by the transfer rate of the telecommunications facility. The system´s rates are limited by the speed of the terminal, by the transfer rate, and by the ability of the computer to process data for output as rapidly as it can be accepted. (Signaling speed includes the start bit, 1 or 2 stop bits and the parity bit, while user and system rates include only the 7 information bits.)

	Speed in Bits per Second			
	Signaling Speed	User	System	Average of both
TYMNET (ref. 5)	110.	3.5	35.0	-
GE Information Services (ref. 5)	110.	-	49.	-
	300.	-	147.	-
Jackson, Stubbs and Fuchs of Bell Telephone Labs (ref. 6 & 7):				
moderately loaded scientific	110.	3.4	61.6	25.2
heavily loaded scientific	110.	1.9	14.7	10.5
moderately loaded business	150.	5.6	58.1	28.0

It may be concluded from this data that terminals and computer systems do not have the same transfer rate requirements. People type much more slowly than computer systems, so that an asymmetric communications link (with respect to transfer rate) may be appropriate for interactive computing.

Even on computer output, the average transfer rate did not approach the maximum achievable. Since computer output over a communication line usually does take place at the maximum achievable rate, we may conclude that computer output is rather bursty. That is, high data rate transfers are separated by lengthy intervals when no data is transferred. It would seem to be desirable to utilize some form of multiplexing technique, in order to achieve higher line utilization.

In view of these conclusions, transfer rate does not appear to be the limiting variable for interactive computer applications. However, it is true that interactive systems in use today were designed for the rather low transfer rates that were commonly available when they were designed. Furthermore, interactive applications generally require much higher burst transfer rates than average transfer rates. In such systems, excess line capacity is frequently provided to accommodate these burst throughput requirements during peak periods.

For RJE (remote job entry) terminals, the transfer rates needed are limited by the speed of the input/output devices used. The following are typical maximum transfer rates:

	Maximum transfer rate in bits per second
Line printer: 120 characters per line 300 lines per minute	4200.
Card reader: 80 characters per card 600 cards per minute	5600.
Card punch: 80 characters per card 91 cards per minute	840.

For computer-to-computer traffic, transfer rates approaching channel speeds for local peripherals (such as, magnetic tape units or rotating storage) are desirable. Typical speeds for such devices are in the range of one million bps and up, though such transfers tend to be rather bursty. Most existing computer-to-computer traffic operates at much lower data rates, with 50 Kbps being the highest rate commonly in use today.

1.6 Typical Transfer Rates Provided by Telecommunications Facilities

Most common carrier offerings utilize analog circuits. These consist of subvoice-grade channels (for teletype-speed devices and other very low speed signaling), voice-grade channels, and wideband channels. The sub voice-grade channels provide signaling rates in the 30 to 150 bit per second range. Voice-grade channels provide a nominal three kilohertz channel that permits data signaling rates from 110 to 4800 bits per second (9600 if leased lines are used), depending upon the type of modem used. Wideband group channels currently permit signaling rates to 50 Kbps, with higher data rates being technically feasible.

Recently, communications services based on digital technology have been offered. These have signaling rates in the 2.4 to 56 Kbps range. However, most computer data transmission is still done via voice-grade, analog channels.

Transfer rates for the voice-grade dial network tend to be somewhat lower than for comparable leased lines using the same signaling rates. This is due to the higher error rates caused by the increased impulse noise on the dial network from the switching equipment and, in the case of some modems, from the effect of variations in the amplitude and delay distortion from one connection to the next.

The Bell System's 1969-70 Connection Survey (ref. 8) contains data on several hundred dialed (voice-grade) telephone connections between numerous cities at data signaling rates of from 150 bps to 4800 bps. These are used in the NBS Technical Note 779 (ref. 9) to derive transfer rates using the ANSI X3.44 - 1974 term, TRIB, and the Basic Mode Control procedure of X3.28 - 1971 for signaling rates from 1200 to 4800 bps. (At the time of the Connection Survey, 4800 bps was the practical upper limit for data communication on the dial network, while up to 9600 bps was possible using conditioned leased lines.) For a near optimal block size of 10,000 bits, the following TRIB values were found:

Signaling Rate bps	TRIB 50th percentile	TRIB 90th percentile
1200	800	750
2000	1600	1550
3600	2800	2500
4800	3600	2700

In a packet-switched computer network (such as the ARPANET), the transfer rate for actual user data is likely to be significantly less than the signaling rate for the backbone communications system. In the ARPANET, for example, the internode signaling rate is 50 Kbps. This is degraded to an effective rate of about 40 Kbps by the overhead of the internode control procedures, but experiments have shown that rates as high as 60 Kbps can be achieved because of the opportunity for the parallel transfer of data through alternative routing (ref. 37). However, for bulk file transfers run using the file transfer protocol, rates in the range of 4 to 10 Kbps were experienced by NBS users. These lower rates might be attributed to end-to-end protocol overhead within the communications subnet, end-to-end protocol overhead between hosts, the inability of the heavily loaded hosts to keep up with higher data rates, or other factors.

2. AVAILABILITY

2.1 Definition

The portion of a selected time interval during which the telecommunications facility is capable of performing its assigned function.

2.2 Other Definitions

Availability is defined in ANSI X3.44 - 1974 (for data communication systems using the X3.28 - 1971 character-oriented control procedure) as "The portion of a selected time interval during which the Information Path is capable of performing its assigned data communication function. Availability is expressed as a percentage."

Availability is defined by the Bell System to be the percentage of a long period of time that a channel is operating within its transmission parameter limits (ref. 10).

2.3 Significance of Availability

It is not possible to guarantee 100 percent availability of the telecommunications facility. Thus, an information processing system may lose part or all of its telecommunications service. The significance of this varies with system usage. For some real-time systems, the loss of even part of the telecommunications facility is quite serious. For others, it may be merely an inconvenience, especially if local computing work may be done in the interim.

2.4 Discussion

There are various reasons why a telecommunications facility may be unavailable to the user. Generally these concern equipment failures or an overloading of the facility. However, the concept of availability can also refer to non-temporal factors, such as, geographic coverage, as in the case of some of the newer common carrier offerings, which are initially available only in selected cities. Availability will also be affected if the telecommunications facility is not scheduled for use during certain hours of the day or days of the week.

The users of the switched telephone network may be unable to complete calls due to trunk busy conditions, transmission outages or failures in the switching equipment. Trunk busy conditions occur during times of heavy business usage of the facility, when it may require numerous attempts to establish the desired connection. Transmission outages and failures in the switching equipment may increase the probability of receiving a trunk busy indication, but do not normally prevent all use of the facility, due to the existence of alternative routings that may be automatically used by the telecommunications facility. Total unavailability of the user's lines is relatively rare due to the passive nature of the local user loops and to the use of alternative power sources by most common carriers during power black outs.

Leased line users are not concerned with trunk busy problems, but do have a very real problem with circuit failures. Since their circuits are dedicated, the failure of an electronic component anywhere in the circuit is likely to cause total unavailability of the circuit. The repair or substitution of alternative circuits may require a few hours, so users with the need for a high availability will usually provide access to the switched network or some alternative channel. In the case of using the switched network, there may be some degradation of the performance due to the somewhat lower signaling rates possible on the switched network. Also, since leased and switched lines would normally have local loops in the same multi-wire cable, a local loop problem might well affect both services.

Since modern computer communications systems employ components beyond the basic communications circuits provided by the common carriers, there are more sources of problems. Failure of the node computer (switching computer), the interface to the computer, or a software failure in a host computer, in addition to simple line failure, may make the system unavailable. Also, failures can be either hardware or software in nature, thus complicating the problem of diagnosis and repair.

The availability characteristics of modern packet switching systems is strongly affected by the way in which flow control mechanisms are implemented. Packet switching systems commonly refuse to accept input at times in order to prevent overload, thus denying service to some users while others are being serviced. Such refusals may be relatively short term, such as the refusal to accept data for an already established conversations until buffers are cleared, or relatively longer term, such as could result from restrictions on the number of (virtual) connections permitted between particular points on the network.

The principal approach to be employed in planned data networks for maintaining the availability of the complex system (with many different components, all subject to individual failure) is redundancy. A partial degree of redundancy exists in some current networks, which provide alternate routing of messages. However, this only protects against the failure of one part of the system: the communication circuit itself. Planned networks are including redundant node computers and redundant links from the nodes to the host computers in order to provide full redundancy for the entire path from originator to recipient.

The concept of partial availability may be applied to large networks, since failures may occur which prevent certain communications paths from being used, while others are not affected at all.

2.5 Typical Availability Requirements of Information Processing Systems

The National Library of Medicine (ref. 11) specified an availability of no worse than 95 percent non-busy during peak hours of operation as the requirement for their bibliographic retrieval system. Since the availability of computers is, typically, in the range of from 0.95 to 0.99, this is a fairly typical requirement. However, for some uses this availability is not adequate. Applications, such as, airline reservations or stock brokerage order entry simply cannot tolerate system unavailability. Such installations often use duplexed computers with an availability that may be as high as 0.996. In these cases, the availability of the telecommunications plant should, if possible, be raised to a similiarly high value.

2.6 Typical Availability Figures for Telecommunications Facilities

Relatively little has been published regarding the availability of voice-grade leased lines. However, customer-reported outages provide a source of relevant information. It is not possible to determine from these figures how long the line was out prior to the customer observing the outage and reporting it to the telephone company. Also, short outages are less likely to be reported than long outages. Intrastate and local lines are likely to have different outage characteristics than interstate lines. Loss of availability due to transmission parameters being out of limits is likely to be more of a problem at the higher signaling rates.

The Bell System reports the following customer-reported outages for leased two-point interstate data service. Since the service includes the use of Bell-supplied modems, the availability of the leased circuits themselves is somewhat higher than the figures shown. 72 percent per month with no reported outage; the average number of reported outages per channel per month is 0.56; the average outage time per channel per month is 68 minutes; where at least one outage does occur, the average outage time per channel per month is 241 minutes; and the percentage of outages exceeding 2.0 hours is 23 percent (ref. 10).

The probability of users of the switched telephone network not receiving a trunk busy signal during the peak busy hour has been estimated at 0.94, with 90 percent confidence intervals of plus or minus 0.02 (ref. 12).

AT & T cites a preliminary design objective of an average of at least 99.96 percent availability for its new Dataphone Digital Service (ref.13).

Datran (Data Transmission Company) cites an average availability of all customers' circuits on their new digital data network of 0.9879 for the time interval from April 1, 1974 to July 24, 1974. Since most of the outages involved subscriber loop problems that were being corrected, the availability was said to be improving drastically (ref. 14). Datran claims that the availability of the high speed channels over the "backbone" of their network (thus, excluding customer loops which are often leased from other carriers) is 0.9998 (ref. 15).

Early availability statistics for the ARPANET (ref. 16) showed backbone circuit outages that had a weighted average of 1.64 percent down time (98.36 percent availability) for the calendar year 1971, though individual lines varied from zero to 29.43 percent. Statistics for June 1972 through December 1972 for the switching computers (IMPS) showed that these processors were out an average of 1.39 percent of the time due to hardware or software failures and 2.42 percent overall (includes preventative maintenance, site environmental problems, etc.). Again, the variability of the performance of individual machines was large, ranging from zero to 8.69 percent for hardware and software failures and from zero to 9.37 percent overall. Mean time between failure (MTBF) was 228 hours overall, and 411 hours for hardware/software failures. Mean time to repair (MTTR) was 5 hours 29 minutes overall and 5 hours 41 minutes for hardware/software failures.

Recent availability statistics for the ARPANET (ref. 17) show a significant improvement. For the first 8 months of 1974, circuit outages had an average of 0.38 percent (99.62 percent availability). For the same time period, the switching computers were out an average of 0.80 percent due to hardware and software failures and 1.33 percent overall . The MTBF was 403 hours and the MTTR was 3.02 hours for the hardware and software failures. While these data are not necessarily valid for other computer networks and the level of availability is achieved at a relatively high cost, they are indicative of what can be done. The improvements in ARPANET's performance are indicative of similar improvements with new systems which take some time to reach their best level of performance.

3. RELIABILITY

3.1 Definition

The ability to maintain the use of the telecommunications facility until the information transfer has been successfully completed.

3.2 Other Definitions

Reliability is defined in ANSI X3.12 - 1970 (ref. 18) as "The probability that a device will perform without failure for a specified time period or amount of usage."

3.3 Significance of Reliability

The concept of reliability is somewhat similar to that of availability. The failure of any component which prevents messages from being sent from an originator to a recipient affects availability if the system is not in use at the time of the failure and reliability if it is in use. The effect of the failure, however, is the same in either case.

Reliability is distinct from availability, however, in that it describes the performance of a system after it has accepted traffic for delivery. A system which refuses input due to malfunctions may have poor availability characteristics, but a quite acceptable level of reliability.

The significance of the reliability of the telecommunications facility to the information processing system is related to the purpose for which the system is used and to the system's own reliability. For some uses, the failure of the telecommunications facility during the information transfer process is a fairly minor inconvenience, provided that the facility is able to re-establish the communication. (If this failure is due to a condition that prevents further use of the telecommunications facility, then the situation also becomes a matter of availability.)

However, for applications having a time urgency, for lengthy transactions requiring the user to repeat work already done, or for frequent failures, poor reliability is quite significant. To some extent, system design can compensate for the effects of poor reliability, but the effect of such failures can be very significant if they are a common occurance.

Poor reliability can also affect the cost of using a particular service. A high failure rate implies high maintenance costs, which, directly or indirectly, must be paid by the user.

3.4 Discussion

There are various reasons why telecommunications facilities have less than perfect reliability. The causes may be grouped into two classes: 1) those involving the temporary disruption of the information transfer process, and 2) those involving a longer disruption of service, but where a readily available alternative telecommunications facility may be used. (There is also the case where there is a long disruption of service without readily available alternative service, but this becomes more a matter of availability.) Generally, reliability is much less of a problem for most users than availability.

Thus, the approach taken in large networks with respect to reliability is the same as that for availability: redundancy of components to assure successful operation. Large networks also generally provide sophisticated diagnostic capabilities and repair procedures in order to minimize the mean time to repair (MTTR).

3.5 Typical Reliability Requirements of Information Processing Systems

There is no simple numerical expression which can be given for the reliability requirements of information processing systems, as systems vary widely in their response to interruptions of service due to failure of the telecommunications facility. Some systems suspend the disconnected program for a fixed or variable length of time and permit the user to resume processing when communications have been restored. At the other extreme, some systems simply abort the job, causing the processing, and possibly even the files, to be lost.

Frequently, the applications on the information processing system will dictate the reliability requirements. For each particular application, it might be possible to determine a numerical requirement based on the length of time to complete a transaction. Applications, such as, stock brokerage or airlines reservations, which are inherently real-time, will have much more stringent requirements than, for example, interactive program debugging.

3.6 Typical Reliability Figures for Telecommunications Facilities

For the dial telephone network the probability of being able to complete a ten minute call without being erroneously disconnected by the switching network is estimated to be better than 0.998 (ref. 12). The reliability of switched or leased telecommunications channels tends to vary with the length and routing of the circuit through the telecommunications facility, with reliability being reduced somewhat as the number of circuit components is increased.

4. ACCURACY

4.1 Definition

The correctness and completeness of the information accepted by the receiving terminal.

4.2 Other Definitions

Accuracy is defined in ANSI X3.44 - 1974 (for data communication systems using the X3.28 - 1971 character-oriented control procedure) in terms of Residual Error Rate (RER). RER is defined as "the ratio of the sum of (1) erroneous Information Characters accepted by the receiving Terminal Configuration and (2) Information Characters transmitted by the sending Terminal Configuration but not delivered to the receiving Terminal Configuration, to the total number of Information Characters contained in the source data. Within (1) above, erroneous Information Characters may be either undetected, or detected but uncorrected, within the Information Path." (The capitalized terms have specific meanings that are defined in the standard.)

This is computed:

RER = (Ce + Cu + Cd) divided by Ct, where

Ce = Erroneous Information Characters accepted by the receiving Terminal Configuration.

Cu = Information Characters transmitted and assumed by the sending Terminal Configuration to be accepted by the receiving Terminal Configuration (undelivered).

Cd = Information Characters accepted in duplicate by the receiving Terminal Configuration which were not intended for duplication (duplicates).

Ct = Information Characters contained in the source data.

Accuracy is defined in the Office of Telecommunications report OTR 75-54 "Performance Criteria for Digital Data Networks" (ref. 19) as having at least four accuracy performance parameters. These are designated as bit error ratio, false bit ratio, erasure bit ratio and duplicate bit ratio. These are defined in that report as:

Bit Error Ratio - the ratio of the number of incorrect bits in non-duplicated messages, accepted by the user at a correct destination, to the total number of user originated information bits.

False Bit Ratio - the ratio of the number of correct and incorrect bits accepted by the user at incorrect destinations, to the total number of user originated information bits.

Erasure Bit Ratio - the ratio of the number of bits which are not accepted by the user at a correct destination to the total number of user originated information bits.

Duplicate Bit Ratio - the ratio of the number of correct and incorrect but duplicated bits, accepted by the user at a correct destination, to the total number of user originated information bits.

4.3 Significance of Accuracy

The significance of undetected errors varies greatly with the usage of the information processing system. For example, systems used for banking and electronic funds transfer require undetected error rates vastly better than those required by interactive terminals used for bibliographic retrieval, where the high redundancy of the English language reduces the significance of errors. However, inaccurate systems tend to have reduced transfer rates, due to the need to retransmit information which is received with detected errors.

4.4 Discussion

Errors on telecommunications circuits tend to be in error bursts or groups, rather than being randomly distributed, since the causes for the errors tend to be conditions that are present for a time period covering the transmission of several bits. For example, impulse noise often occurs as a series of noise spikes of several milliseconds duration each. This grouping of errors tends to reduce the effectiveness of error detection codes below that for randomly occurring errors.

Early control procedures up through the Basic Mode Control Procedure of X3.28 - 1971 and the IBM Bisync procedure use a combination of vertical and longitudinal parity checking that is reasonably effective for most users.

More recently, polynomial error detection codes have been developed that have much lower undetected error rates and a reduced number of overhead bits required for the error detection process. These are used with the newer control procedures and are placed at the end of the transmission block. It is possible to reduce the undetected error rate to any arbitrarily low value by using codes of greater length. While this is true for other types of codes as well, polynomial codes are particularly efficient in this regard.

4.5 Typical Accuracy Requirements of Information Processing Systems

The accuracy requirements of information processing systems are often not precisely known. The system may have programs for widely different uses, each of which has its own inherent accuracy requirements. Even the individual user may not be able to assess the value of accuracy or the cost of inaccuracy. Also, the effect of undetected errors may involve intangibles, such as, customer dissatisfaction. In a study done for the U. S. Postal Service, the consultant recommended a desired bit error probability of one in 10**12 (ref. 20). In general, the error detection codes used are sufficiently effective relative to the error rates encountered on most telecommunications facilities that the accuracy requirements of most users are satisfied.

4.6 Typical Accuracy Figures for Telecommunications Facilities

Accuracy using vertical and longitudinal parity checking is dependent on the character size (bits per vertical parity bit), block length (bits per longitudinal parity bit) and the error burst characteristics of the telecommunications circuit used. For an error to be undetected, there must be an even number of error bits in each character and an even number of error bits in each bit position of the characters (per longitudinal parity bit). In the simplest case, there would be the same two bits in error in each of two characters.

Due to the error burst tendencies of most telecommunications circuits, it is difficult to obtain good estimates of the undetected error rate using vertical and longitudinal parity checking. According to Martin (ref. 21), the residual error rate on switched telephone facilities having an error rate of one in 10**5 (in this report double asterisks denote exponentiation) might be in the range of one bit error in 10**7 to one in 10**9 bits. Since there is one parity bit for every seven information bits plus a parity bit in every track, the efficiency of this error detection process is less than 0.875.

With the polynomial error detection codes, both the undetected error rate and the efficiency can be improved. The undetected error rate obtained with a polynomial code is dependent on the length of the code, with the longer codes providing vastly better detection capabilities, and on the error burst characteristics of the telecommunications circuit used. The detection capabilities of polynomial codes are defined by Peterson (ref. 22 and 23). The efficiency of polynomial codes is quite good. A 16 bit code used with a block length of 1000 bits has an efficiency of 98.4 percent, while providing very effective checking of the data.

A properly chosen polynomial code will detect 100 percent of: single bits in error, two bits in error, an odd number of bits in error, and all error bursts equal to or less than the number of code bits. It will also detect a very high percentage of longer bursts up to the length of the code. The length of a polynomial code is defined to be 2**(m-1), where m is the number of code bits used.

Using theorem 8.3 (page 152) of Peterson's book (ref. 22), a 16 bit polynomial code is shown to be capable of detecting all errors due to error bursts less than 17 bits long and to have an undetected error probability of 2**-15 for 17 bit error bursts and 2**-16 for bursts that are more than 17 bits long, up to the code length of 32,768 bits. (Other considerations limit block lengths to much less than 32,000 bits.)

To relate this to a typical system application, error burst statistics are needed. Using the Bell System survey for dial telephone connections (ref. 8), the 50th percentile connection for a 2000 bits per second modem was found to have an error burst rate of about 5. x 10**-7. Using the Burst Length Distributions chart from the same report, about one percent of the error bursts were found to be 17 bits long and about 14 percent were found to be longer than 17 bits.

Thus, the theoretical probability of an undetected error in this particular situation would be about:

U = (5. x 10**-7)(0.01 x 3.1 + 0.14 x 1.5)(10**-5)

or about one bit in 10**12. At 2000 bits per second this would mean more than 138,000 hours (15.7 years) between undetected errors.

Packet switched networks with much greater transmission rates tend to use somewhat longer polynomial error detection codes to achieve even greater accuracies. For example, the ARPANET uses a 24-bit polynomial code, for which an undetected bit error rate of one in 10**12 is claimed, or about one undetected error per year in the network (ref. 35).

The Bell System supports a leased line minimum performance of one 1000-bit block in error per 100 blocks transmitted when certain Bell synchronous modems are used and when some other specified conditions are met. (Support indicates a performance goal of the Bell System that is not included in the tariff as being legally obligatory.) This applies to the Bell 201A modem at a signaling rate of 2000 bps; the Bell

201C at 2400 bps; and the Bell 208A at 4800 bps. The Bell System also supports a minimum performance of 1 bit error per 100,000 bits transmitted when certain Bell asynchronous modems are used and some other specified conditions are met. This applies to the Bell 202C, D, E and R modems at 1200, 1400 and 1800 bps for basic, C1 and C2 channel conditioning, respectively. At least 80 percent of all leased lines should meet these performance goals (ref. 10).

The digital data offerings of the common carriers have defined accuracy in terms of average error-free seconds. Datran guarantees 99.95 percent average error-free seconds during customer transmissions.

Datran also cites an error performance measured on a circuit from Dallas, Texas to St. Louis, Missouri for a six day period in July 1974 as providing an "extrapolated probability of error free seconds" of 99.98 percent. They state that this level of service will be provided on switched circuits, since there should be virtually no additional errors introduced by the switch itself (ref. 14).

AT & T cites a preliminary design objective of 99.5 percent error-free seconds at 56 Kbps and better performance at the lower rates of 9.6, 4.8 and 2.4 Kbps for their Dataphone Digital Service (ref. 13).

5. CHANNEL ESTABLISHMENT TIME

5.1 Definition

The time required for a switched telecommunications facility to provide a circuit connection between a calling terminal and a called terminal.

5.2 Other Definitions

Channel establishment time is contained within the Transfer Overhead Time (TOT) term of ANSI X3.44 - 1974 for data communication systems using the X3.28 - 1971 character-oriented control procedure. Transfer Overhead Time is defined as "A measure of the individual communication phase times required to establish and dis-establish an Information Path in order to effect an information transfer."

5.3 Significance of Channel Establishment Time

Channel establishment time includes both the time to operate the dialing mechanism and the time for the network to complete the connection. This time is probably not too significant to most people who are using the network for its intended purpose of making voice calls. Remote terminal users that have lengthy dialogues with the computer may not find this time too significant either, especially if the sign-on procedure is lengthy. However, a terminal user who must make a great many brief accesses to the computer each day could find the channel establishment time to be quite an annoyance.

In the case of computer-initiated calls (using an automatic calling unit), the several seconds required for switching time, does impose a quite significant delay before the computer can begin its transaction with the remote terminal. While the computer is not idle during this time, it may be a significant time factor to the system. In general, if this time is commensurate with the time required for the data transfer to take place, then there is at least the implication of a significant time delay.

For some information processing applications, the channel establishment time and its fixed cost is the most significant factor in the evaluation of a communications system. Applications that must perform continuous polling of remote terminals tend to be used with leased communication lines that are routed to each terminal, thereby reducing the channel establishment time to that required for modem re-equalization. However, the availability of switched telecommunications service with minimal channel establishment time and common carrier charges computed on a measured time basis with a minimum time of no more than a few seconds (e.g. Datran's Datadial service) will probably cause many users to change from leased line service to switched service.

5.4 Discussion

The concept of channel establishment time as a one-time overhead delay to set up a transmission path for the duration of the communication does not apply for the new data networks based on packet switching technology. In these cases, all messages are delivered with equal average rapidity from sender to receiver. In one view, there is no channel establishment delay; in the other view, this delay is imposed on every communication between sender and receiver, not just the first. However, the delay, as measured for the physical task of establishing a communication path, is considerably less for packet switched networks than for the present circuit switched networks.

Note, however, that for networks of computers there may be a different sort of channel establishment delay introduced by the software in the receiving computer which must condition that computer to communicate according to specifications indicated by the sender. This process of establishing communications on mutually agreeable terms may take several data exchanges in each direction before all procedural matters are resolved.

5.5 Typical Channel Establishment Time Requirements of Information Processing Systems

While not generally stated in system specifications, a terminal user should not have to wait for more than several seconds for the channel to be established.

5.6 Typical Channel Establishment Times Provided by Telecommunications Facilities

For telephone calls to subscribers served by the same telephone end office, the channel establishment time is determined by three factors: 1) the number of digits required; 2) the type of dialing mechanism (rotary or multi-tone pushbutton); and 3) the telephone end office switching machine type. For a typical requirement of 7 digits, rotary dialing time is about 10 seconds and multi-tone pushbuttons about 4 seconds. An automatic (computer-controlled) calling unit needs about the same time, as the fixed dial tone time delay tends to cancel any speed advantage in the dialing process. Telephone office switching time is generally 2 seconds or less.

For telephone calls to subscribers not served by the same end office, the channel establishment time includes about 50 percent more dialing time (if an area code is required) and the time for switching between the various telephone offices and centers. The telephone network is a structured hierarchy with 5 ranks or classes of switching centers. The ranks proceed from end office through toll center, primary center, and sectional center to regional center. There are numerous trunks between offices and centers. Calls entering the network are always routed over the most direct available cross trunk to the destination. However, it is possible for a call to be routed up through the hierarchy to the regional office, across to the regional office for the destination, and down through its hierarchy. In this case, about 25 seconds would be added to the channel establishment time. This is a rare occurrence even for transcontinental calls (ref. 24).

Long distance calls typically have 3 to 5 trunks. The average connect time for calls up to 180 miles is 11.1 seconds; for 180 to 725 miles it is 15.6 seconds; and for 725 miles to 2900 miles it is 17.6 seconds. (The standard deviations are 4.6, 5.0 and 6.6 seconds, respectively. Ref. 12.)

Datran cites a switched network response time of less than 3 seconds 99 percent of the time and a mean response time of less than 1 second (ref. 14).

6. NETWORK DELAY

6.1 Definition

The time required for information to travel through the telecommunications channel to the receiving terminal or computer.

6.2 Other Definitions

The ANSI X3S3 Task Group 5 is considering the following definition: Message Transfer Time (MTT) "is the time required for a message to be transmitted from a source and accepted at the designated sink".

6.3 Significance of Network Delay

Depending upon the capabilities of the terminal facilities (particularly with respect to buffer capacity), the application's information interchange requirements (volume), the response characteristics (real-time or other), the mode of channel operation (half or full-duplex), and the reliability of the channel, the delay of the channel can be of varying degrees of significance. For example, an interactive application involving short bursts of conversational information interchange, operated in echoplex mode (full-duplex with echo checking of the input) requires a minimum network delay. Otherwise, the operator's striking of the keys and the data input process must be retarded by the character echo response time. Similarly, large network delays can seriously impede the information interchange process when substantial quantities of data must be transferred in a half-duplex channel. In the case of very long delays, the system may have to provide buffer storage for two or more blocks of information, as the negative acknowledgement may not be received until one or more additional blocks have been sent.

6.4 Discussion

Telephone circuit delays are usually not very significant. However, when satellite links are used, the network delay becomes much longer and can be a problem. In the case of full duplex echoing, a delay of a few hundred milliseconds can be quite annoying for terminal users. In the case of half duplex operation, long network delays will seriously reduce the transfer rate with the shorter block lengths (due to the greatly increased time required for acknowledgements) and will tend to force the system to use a longer block length than would otherwise be optimal. Even with optimally chosen block lengths, propagation delay tends to reduce the transfer rate somewhat, as the longer block lengths will increase the probability of a block containing an error, thereby, increasing the percentage of blocks that must be retransmitted.

Message and packet switched communication systems introduce additional network delay because of the time spent processing the message or packet at each node through which it passes. In the case of a packet-switched system, which breaks up the longer messages into packets for transmission through the network and then re-assembles them prior to delivery, any time spent in message dis-assembly/re-assembly (including time spent waiting for constituent packets at the re-assembly point) is part of the network delay. Packet-switched systems with adaptive routing are affected not only by the network delay, but by variations in that delay, as the re-assembly of the packets is affected by such variations. This may become critical in extreme cases. Thus it may be necessary to specify the standard deviation of the network delay in these cases.

If the definition of network delay suggested above is to be used for computer networks, the time interval of concern must be carefully identified. Is it from the time of transmission for the first character of the input to the receipt of that character on the output, or is the last character a better starting point? For packet switched systems, the first-to-first measure might be unrealistic, if intermediate packets in a message were subject to internal delays. Measuring from last to last may not be fair either. If the input to the system was large and quite bursty, the system may have had the opportunity to forward the first parts of the message before the last parts were entered. Thus, it is necessary to compensate for irregularities in entering data, perhaps by taking an average of the two types of measures. It may also be desirable to adjust for the delays in data entry or exit, as these may or may not be considered to be a part of the network delay.

6.5 Typical Network Delay Requirements of Information Processing Systems

The bibliographic retrieval system of the National Library of Medicine (ref. 11) specified the round trip network delay as not being in excess of 2.0 seconds. (This requirement was due to the terminal user's need for a reasonable response time.)

For applications requiring echoing of the input data, the overall response (network delay in both directions, line turnaround delay and computer response) should not exceed 0.1 to 0.2 seconds. For applications involving users engaged in high-intensity "brainstorming", requiring the ready access of data from his own "short-term memory", the response delay should not exceed 2 seconds. However, some uses may be able to tolerate response delays of up to 15 seconds or more (ref. 25).

The goal for the ARPANET (ref. 26) was a maximum round trip delay of 0.5 seconds for the transmission of a 1000-bit packet.

6.6 Typical Network Delays Provided by Telecommunications Facilities

Terrestrial telephone circuit network delays are typically in the range of from 2 to 15 milliseconds. When satellite circuits are used the delays are typically 300 milliseconds long. ARPANET network delays vary with the queueing delays at each node computer and with the number of node computers through which the messages must pass. Delays are in the 20 to 120 millisecond range (ref. 26).

7. LINE TURNAROUND DELAY

7.1 Definition

The time delay required by half-duplex circuits to reverse the direction of transmission.

7.2 Significance

The line turnaround delay reduces the transfer rate of half-duplex circuits by increasing the time required for acknowledgements. It may necessitate a longer block length than would otherwise be optimal, since turnaround delay, being a fixed constant, is less significant for longer block lengths. (This delay does not apply to full-duplex circuits, since telecommunications channels are permanently established in both directions.)

7.3 Discussion

Leased telephone circuits are usually available as four-wire, full -duplex circuits. Dial telephone circuits are normally two-wire, unless the user is willing to double the line costs by using two telephone circuits. At the lower signaling rates, the modems usually are designed for two-way simultaneous operation by using half of the available bandwidth for transmission in each direction. However, at the higher signaling rates this is not possible and most high speed modems used on the dial network use two-way alternate operation. (There are some modems that have a very slow speed reverse channel by using a small portion of the bandwidth, but they have their own performance problem with the time required for sending the acknowledgement message.) In two-way alternate operation the echo suppressors must either be disabled at all times or they must be reversed each time that the direction of transmission is reversed. When echo suppressors are not used, time must be allowed for echos to subside before transmitting in the reverse direction.

7.4 Typical Line Turnaround Delay Requirements of Information Processing Systems

Line turnaround delay is more significant for short block length transmissions than for longer block lengths. For systems operating asynchronously in "echo-plex" mode, where transmission is a character at a time and all characters are echoed for verification and/or printing, very short turnaround delays are essential. For longer block lengths, somewhat longer turnaround delays are tolerable.

The turnaround delay requirements will also be closely coupled to the propagation delay requirements described in section 6.5.

7.5 Typical Line Turnaround Delays Provided by Telecommunications Facilities

The time required for reversing the echo suppressors varies somewhat, so modems that do not disable the echo suppressors should be adjusted for the longest reversal delay that they are likely to encounter. A time of 150 milliseconds is usually considered to be safe. On short distance circuits, the common carrier may omit the echo suppressors. In this case, a delay of 50 milliseconds is generally adequate for the echos to subside.

Modems that disable the echo supressors may still have substantial line turnaround delay, as high speed modems also have a minimum turnaround delay required for re-training the equalizer. Modern modems typically need 40 to 50 milliseconds to this.

Systems that cannot tolerate these turnaround times can generally specify full-duplex circuits, thereby, avoiding the delay altogether. Modern packet networks, for example, all specify full-duplex circuits.

8. TRANSPARENCY

8.1 Definition

The absence of code or procedural constraints imposed by the telecommunication facility.

8.2 Other Definitions

The IBM Corporation in manual GA27-3093-0, "IBM Synchronous Data Link Control, General Information", defines transparency as follows: "In data communications, the quality of information data when there is no possibility of interference with data link control, regardless of format or content. The transparent information is unrecognized by data link controls."

8.3 Significance of Transparency

Transparency is a qualitative term that refers to the absence of code or procedural constraints being imposed on the information processing system. Its significance is relative to the degree that system changes or user education are required to compensate for lack of transparency by the telecommunications plant.

8.4 Discussion

Code transparency may be impaired by the modem's inability to transmit long strings of ones or long strings of zeros. It may also be impaired if certain bit code combinations are not permitted. For example,if the Basic Mode Control Procedure of X3.28 - 1971 is used, then an EOT (end of transmission) code cannot be used within the text of a message, as it may cause a disconnect.

In the case where the link control procedure is part of the telecommunications service provided, additional transparency problems exist. Frequently, control procedure transparency is achieved by modifying the user's text before and after transmission in order to remove any illegal bit sequences. This approach is taken both with the character-oriented and the newer bit-oriented link control procedures, and may be accomplished either in software or hardware. The result of the modification, however, is a longer message than was originally specified to be sent. In a worst case situation the message actually transmitted might be nearly twice as long as the originally specified message. Thus, poor transparency characteristics for a control procedure can be overcome, but at the expense of some reduction in the transfer rate.

Transparency is a complex concept when applied to a network of computers. Packet networks undertake to deliver data as supplied to them; in that sense the network is transparent, since the data put in is delivered without examination or alteration. At a higher level, computer networks impose severe procedural restrictions on their usage. Protocol conventions are usually established for both terminal users and computer programs that communicate via the network. Thus, present computer networks would have to be judged to be quite opaque.

8.5 Typical Transparency Requirements of Information Processing Systems

The bibliographic retrieval system of the National Library of Medicine (ref. 11) specified that "The proposed configuration must in no way require that NLM applications programs be redesigned to accommodate any peculiarities of such configurations. The configuration must be transparent to all transmissions between the user and the NLM and SDC (System Development Corp.) computers once the link has been established."

8.6 Typical Transparency Characteristics of Telecommunications Facilities

Code transparency is not a problem with modern modems. The transparency characteristics of computer networks were discussed in 8.4

9. SECURITY

9.1 Definition

The ability to protect the information being transferred by the telecommunications facility from deliberate or inadvertent disclosure, modification or loss.

9.2 Significance of Security

The importance of security in an information processing system is dependent directly upon the usage of the system. For some government and military systems, security is considered to be an extremely significant factor. For many commercial systems, the potential threat of industrial espionage or malicious damage is taken very seriously.

9.3 Discussion

Protection of information involves the entire system. This includes the computer, the telecommunications facility, the terminal, the terminal user's authorization to use the system and the protection of user's files from each other. The telecommunications facility is vulnerable to the unauthorized disclosure of information being transmitted through it. For this reason, users who feel that they have security needs are likely to seek methods of improving telecommunications security.

The most effective method of improving telecommunications security is encryption. Encryption is the substitution of other characters or bits for the data characters or bits being transmitted, or the transposition of the data characters or bits within the message for purposes of secrecy. This is done according to a predetermined method, thus permitting the receiving end to decrypt the characters or bits back to the original data. Most encryption methods use a "key", which consists of several characters or bits used to specify the substitution or transposition process to be used. Security is achieved by the secrecy of either the encryption method or the key used.

Packet switched networks may be viewed as providing some minor degree of security by virtue of the alternative routing of successive packets of a message. Thus, eavesdropping on any single circuit may not provide access to the complete message. Encryption techniques may also be used in such networks, both on an end-to-end and link-by-link basis. If end-to-end encryption is used, the message will also be protected even while it is in the switching computers.

Packet and message switching systems can also implement "closed user groups" as an aid to overall security. This facility defines a logical network between a particular group of systems and prevents unauthorized user modes from even establishing connections with those systems.

9.4 Security Requirements of Information Processing Systems

Most systems do not yet impose security requirements on the telecommunications facility. Systems that are felt to have such needs normally achieve the necessary security by means of physical security for the facility and encryption of the data passing through the facility. The National Bureau of Standards has published an encryption algorithm that is being proposed as a Federal Information Processing Standard (ref. 36).

9.5 Security Characteristics Provided by Telecommunications Facilities

Wiretapping of telephone facilities is difficult for outsiders, except where the wiring enters the user's premises. Wiring within the user's premises may be protected by physical means and by periodic inspections. Further protection may be obtained by the use of some form of burglar alarm, possibly monitored by the computer. The complexity of the cable wiring outside of the user's premises and the use of multiplexing equipment within the telephone company premises are generally sufficient to make wiretapping unlikely, unless there is collusion with telephone company personnel (ref. 27).

Packet networks, such as the ARPANET, may be criticized on the grounds that the switching computers reside on customer premises and, thus, could be tampered with by unauthorized personnel. Planned commercial packet switching systems do not intend to operate in this manner, but rather will locate the switching computers off customer premises.

CONCLUSIONS

This report has identified and discussed nine parameters to be used in the evaluation and comparison of alternative data communications services which are used by information processing systems. While this list of parameters is not exhaustive, it is felt to represent the key factors.

The number of alternatives available today for data communications is not small, and it is growing as more and more specialized carriers begin to offer service. Many of these service offerings are quite different, ranging from "raw" communications, as in the leasing of a circuit, up through "packaged" or "value-added" service, where the vendor undertakes to provide an improved communications service above mere circuit capacity.

The parameters which have been described enable a prospective user to compare these diverse offerings. Differences between services will show up in the values of the various parameters, and the user will be able to identify what is being purchased in terms of specific performance characteristics. For example, the difference in price between circuit leasing and value-added packet-switched service should show up in improved availability, accuracy and channel establishment time. With differences identified, the user will be able to decide, along with the price of the each of the different offerings, which service is best suited for the particular application.

The task of selection is not an easy one, however. Users must exercise considerable caution in comparing data from different services. It is particularly important to be sure that the data are expressed in the same way and based on comparable sets of assumptions. For example, certain parameters may vary according to the load on the system or the peak to average ratios might be quite different. Users should seek to obtain these parameters for the same conditions under which they would expect to use the system.

One entire area which has not been discussed in this report, but which may be very important in the selection between alternatives is that of network management. The inclusion of such capabilities as a centralized trouble shooting desk, regional support personnel, sophisticated diagnostic routines, and complete statistical reports showing usage, traffic patterns and load, may be quite important to prospective users. Of course, many of these factors will show up in the various performance measures, such as, availability, or, on the other side, in the cost of the service. Users must decide for themselves the value of any management services which do not directly affect performance.

ACKNOWLEDGEMENTS

The authors would like to express their appreciation to Dr. Peter McManamon of the Office of Telecommunications and to Mr. George Sugar of the National Bureau of Standards for their help in numerous discussions. Also, to Mr. William Combs of Tymshare Corp., Dr. Dixon Doll of DMW Telecommunications Corp., Mr. Lynn Hopewell of Network Analysis Corp., Dr. Barry Wessler of Telenet Communications Corp. and Dr. Dennis Branstad of NBS for their assistance in reviewing the manuscript for this report.

BIBLIOGRAPHY

(1) American National Standard X3.44 - 1974, "Determination of Performance of Data Communications Systems".

(2) American National Standard X3.28 - 1971, "Procedures for the Use of the Communication Control Characters of the American National Standard Code for Information Interchange in Specified Data Communication Links".

(3) American National Standard X3.1 - 1969, "Signaling Speeds for Data Transmission".

(4) Blanc, R. P., "Review of Computer Networking Technology", NBS Technical Note 804, Jan. 1974.

(5) Schwartz, M., Boorstyn, R.,and Pickholtz, R., "Terminal-Oriented Computer Communications Networks", Proc. IEEE, Nov. 1972, pp. 1408-1422.

(6) Jackson, P. and Stubbs, C., "A Study of Multi-Access Computer Communications", SJCC, 1969, pp.491-504.

(7) Fuchs, E. and Jackson P., "Estimating Distributions of Random Variables for Certain Computer Communications Traffic Modes", CACM, Vol. 13, No. 12, Dec. 1972, pp. 752-757.

(8) "1969-70 Switched Telecommunications Network Connection Survey", Bell System Technical Reference PUB 41007, April 1971.

(9) Grubb, D. S., "Data Communications System Throughput Performance Using High Speed Terminals on the Dial Telephone Network", NBS Technical Note 779, May 1973.

(10) "Data Communications Using Voiceband Private Line Channels", Bell System Technical Reference PUB 41004, Oct. 1973.

(11) "Request for Proposal NLM 72-102 ADP/Telecommuncations Support for a Remote-Access Bibliographic Information Retrieval System", National Library of Medicine, 1972.

(12) "Data Communications Using the Switched Telecommunications Network", Bell Systems Technical Reference PUB 41005, May 1971.

(13) "Digital Data System Channel Interference Specifications", Bell System Technical Reference PUB 41021, March 1973.

(14) Farinholt, E. V., "Datran's Switched Digital Network", IEEE Northeast Electronics Research and Engineering Meeting, Boston, Mass., Oct. 1974.

(15) "High Speed Private Channel Data Service", Data Transmission Company.

(16) Network Analysis Corp., "ARPANET: Design, Operations, Management and Performance", Report to Rome Air Development Center, April 1973.

(17) Bolt, Beranek and Newman Inc., report to the Advanced Research Projects Agency, Sept. 1974.

(18) American National Standard X3.12 - 1970, "Vocabulary for Information Processing".

(19) McManamon, P., Rosich, R., Payne, J. and Miles, M., "Performance Criteria for Digital Data Networks", Office of Telecommunications OTR 75-54, Jan. 1975.

(20) McManamon, P., "Satellite Communications for the U. S. Postal Service", Second Annual Telecommunications for Government Conference, Boulder, Colorado, Oct. 1973.

(21) Martin, J., "Teleprocesing Network Organization", Prentice-Hall, 1970.

(22) Peterson, W. W., "Error-Correcting Codes", Wiley, 1961.

(23) Peterson, W. W. and Welden, E. J., "Error-Correcting Codes", MIT Press, 1972.

(24) AT & T Long Lines Division, information provided to the Federal Telecommunications Standards Committee, June 1974.

(25) Miller, R. B., "Response Time in Man-Computer Conversational Transactions", FJCC 1968, pp. 267-277.

(26) Heart, F., Kahn, R., Ornstein, S., Crowther, W. and Walden, D., "The Interface Message Processor for the ARPA Computer Network", SJCC, 1970, pp. 551-567.

(27) Martin, J., "Security, Accuracy and Privacy in Computer Systems", Prentice-Hall, 1973.

(28) Renninger, C. and Branstad, D., "Government Looks at Privacy and Security in Computer Systems", NBS Technical Note 809, Feb. 1974.

(29) Renninger, C., "Approaches to Privacy and Security in Computer Systems", NBS Special Publication 404, Sept. 1974.

(30) Federal Information Processing Standards Publication 31, "Guidelines for Physical Security and Risk Management", 1974.

(31) Boustead, C. and Mehta, K., "Getting Peak Performance on a Data Channel", Data Communications, July 1974.

(32) Doll, D., "Where to Spend Money to Improve System Availability", Data Communications, Nov. 1974, pp. 57-61.

(33) Frank, H. and Hopewell, L., "Network Reliability", Datamation, August 1974, pp. 85-91.

(34) Martin, J., "Systems Analysis for Data Transmission", Prentice-Hall, 1972.

(35) Roberts, L. G. and Wessler, B. D., "Computer Network Development to Achieve Resource Sharing", Proc. SJCC, 1970, pp. 543-549.

(36) National Bureau of Standards, "Encryption Algorithm for Computer Data Protection", Federal Register, Volume 40, Number 52, March 17, 1975.

(37) Cole, Gerald D., "Performance Measurements on the ARPA Computer Network", Proc., Second Symposium on Problems in the Optimization of Data Communications Systems, October 20-22, 1971, pp. 39-45.

NBS-114A (REV. 7-73)

U.S. DEPT. OF COMM. BIBLIOGRAPHIC DATA SHEET	1. PUBLICATION OR REPORT NO. NBS TN-882	2. Gov't Accession No.	3. Recipient's Accession No.
4. TITLE AND SUBTITLE Criteria for the Performance Evaluation of Data Communications Services for Computer Networks			5. Publication Date September 1975
			6. Performing Organization Code 650.01
7. AUTHOR(S) Dana S. Grubb and Ira W. Cotton			8. Performing Organ. Report No.
9. PERFORMING ORGANIZATION NAME AND ADDRESS NATIONAL BUREAU OF STANDARDS DEPARTMENT OF COMMERCE WASHINGTON, D.C. 20234			10. Project/Task/Work Unit No. Project 6502131
			11. Contract/Grant No. STRS
12. Sponsoring Organization Name and Complete Address *(Street, City, State, ZIP)* Computer Systems Engineering Division Institute for Computer Sciences and Technology National Bureau of Standards Washington, D. C. 20234			13. Type of Report & Period Covered Interim Report
			14. Sponsoring Agency Code

15. SUPPLEMENTARY NOTES

16. ABSTRACT *(A 200-word or less factual summary of most significant information. If document includes a significant bibliography or literature survey, mention it here.)*

In general, when telecommunications services are used as a means of interchanging information between information processing systems, or between terminals and systems, a number of parameters determine how well that interchange is performed. This report examines the following characteristics of telecommunications services:

1. Transfer Rate
2. Availability
3. Reliability
4. Accuracy
5. Channel Establishment Time
6. Network Delay
7. Line Turnaround Delay
8. Transparency
9. Security

These terms are all defined and their significance discussed. The effects of these factors on data communication networks are illustrated.

17. KEY WORDS *(six to twelve entries; alphabetical order; capitalize only the first letter of the first key word unless a proper name; separated by semicolons)*

Computer communications; computer networking; data communications; networks; performance requirements; telecommunications.

18. AVAILABILITY	19. SECURITY CLASS (THIS REPORT)	21. NO. OF PAGES
[X] Unlimited [ ] For Official Distribution. Do Not Release to NTIS	X UNCLASSIFIED	36
[X] Order From Sup. of Doc., U.S. Government Printing Office Washington, D.C. 20402, SD Cat. No. C13. 46:882 [X] Order From National Technical Information Service (NTIS) Springfield, Virginia 22151	20. SECURITY CLASS (THIS PAGE) UNCLASSIFIED	22. Price 85 cents

USCOMM-DC 29042-P74

NBS TECHNICAL PUBLICATIONS

PERIODICALS

JOURNAL OF RESEARCH reports National Bureau of Standards research and development in physics, mathematics, and chemistry. It is published in two sections, available separately:

• **Physics and Chemistry (Section A)**

Papers of interest primarily to scientists working in these fields. This section covers a broad range of physical and chemical research, with major emphasis on standards of physical measurement, fundamental constants, and properties of matter. Issued six times a year. Annual subscription: Domestic, $17.00; Foreign, $21.25.

• **Mathematical Sciences (Section B)**

Studies and compilations designed mainly for the mathematician and theoretical physicist. Topics in mathematical statistics, theory of experiment design, numerical analysis, theoretical physics and chemistry, logical design and programming of computers and computer systems. Short numerical tables. Issued quarterly. Annual subscription: Domestic, $9.00; Foreign, $11.25.

DIMENSIONS/NBS (formerly Technical News Bulletin)—This monthly magazine is published to inform scientists, engineers, businessmen, industry, teachers, students, and consumers of the latest advances in science and technology, with primary emphasis on the work at NBS. The magazine highlights and reviews such issues as energy research, fire protection, building technology, metric conversion, pollution abatement, health and safety, and consumer product performance. In addition, it reports the results of Bureau programs in measurement standards and techniques, properties of matter and materials, engineering standards and services, instrumentation, and automatic data processing.

Annual subscription: Domestic, $9.45; Foreign, $11.85.

NONPERIODICALS

Monographs—Major contributions to the technical literature on various subjects related to the Bureau's scientific and technical activities.

Handbooks—Recommended codes of engineering and industrial practice (including safety codes) developed in cooperation with interested industries, professional organizations, and regulatory bodies.

Special Publications—Include proceedings of conferences sponsored by NBS, NBS annual reports, and other special publications appropriate to this grouping such as wall charts, pocket cards, and bibliographies.

Applied Mathematics Series—Mathematical tables, manuals, and studies of special interest to physicists, engineers, chemists, biologists, mathematicians, computer programmers, and others engaged in scientific and technical work.

National Standard Reference Data Series—Provides quantitative data on the physical and chemical properties of materials, compiled from the world's literature and critically evaluated. Developed under a world-wide program coordinated by NBS. Program under authority of National Standard Data Act (Public Law 90-396).

NOTE: At present the principal publication outlet for these data is the Journal of Physical and Chemical Reference Data (JPCRD) published quarterly for NBS by the American Chemical Society (ACS) and the American Institute of Physics (AIP). Subscriptions, reprints, and supplements available from ACS, 1155 Sixteenth St. N. W., Wash. D. C. 20056.

Building Science Series—Disseminates technical information developed at the Bureau on building materials, components, systems, and whole structures. The series presents research results, test methods, and performance criteria related to the structural and environmental functions and the durability and safety characteristics of building elements and systems.

Technical Notes—Studies or reports which are complete in themselves but restrictive in their treatment of a subject. Analogous to monographs but not so comprehensive in scope or definitive in treatment of the subject area. Often serve as a vehicle for final reports of work performed at NBS under the sponsorship of other government agencies.

Voluntary Product Standards—Developed under procedures published by the Department of Commerce in Part 10, Title 15, of the Code of Federal Regulations. The purpose of the standards is to establish nationally recognized requirements for products, and to provide all concerned interests with a basis for common understanding of the characteristics of the products. NBS administers this program as a supplement to the activities of the private sector standardizing organizations.

Federal Information Processing Standards Publications (FIPS PUBS)—Publications in this series collectively constitute the Federal Information Processing Standards Register. Register serves as the official source of information in the Federal Government regarding standards issued by NBS pursuant to the Federal Property and Administrative Services Act of 1949 as amended, Public Law 89-306 (79 Stat. 1127), and as implemented by Executive Order 11717 (38 FR 12315, dated May 11, 1973) and Part 6 of Title 15 CFR (Code of Federal Regulations).

Consumer Information Series—Practical information, based on NBS research and experience, covering areas of interest to the consumer. Easily understandable language and illustrations provide useful background knowledge for shopping in today's technological marketplace.

NBS Interagency Reports (NBSIR)—A special series of interim or final reports on work performed by NBS for outside sponsors (both government and non-government). In general, initial distribution is handled by the sponsor; public distribution is by the National Technical Information Service (Springfield, Va. 22161) in paper copy or microfiche form.

Order NBS publications (except NBSIR's and Bibliographic Subscription Services) from: Superintendent of Documents, Government Printing Office, Washington, D.C. 20402.

BIBLIOGRAPHIC SUBSCRIPTION SERVICES

The following current-awareness and literature-survey bibliographies are issued periodically by the Bureau:

Cryogenic Data Center Current Awareness Service

A literature survey issued biweekly. Annual subscription: Domestic, $20.00; foreign, $25.00.

Liquefied Natural Gas. A literature survey issued quarterly. Annual subscription: $20.00.

Superconducting Devices and Materials. A literature survey issued quarterly. Annual subscription: $20.00. Send subscription orders and remittances for the preceding bibliographic services to National Technical Information Service, Springfield, Va. 22161.

Electromagnetic Metrology Current Awareness Service Issued monthly. Annual subscription: $100.00 (Special rates for multi-subscriptions). Send subscription order and remittance to Electromagnetics Division, National Bureau of Standards, Boulder, Colo. 80302.