INTERNATIONAL TELECOMMUNICATION UNION

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

I.350

(03/93)

INTEGRATED SERVICES DIGITAL NETWORK (ISDN)

OVERALL NETWORK ASPECTS AND FUNCTIONS

GENERAL ASPECTS OF QUALITY OF SERVICE AND NETWORK PERFORMANCE IN DIGITAL NETWORKS, INCLUDING ISDNs

ITU-T Recommendation I.350

(Previously "CCITT Recommendation")

FOREWORD

The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of the International Telecommunication Union. The ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Conference (WTSC), which meets every four years, established the topics for study by the ITU-T Study Groups which, in their turn, produce Recommendations on these topics.

ITU-T Recommendation I.350 was revised by the ITU-T Study Group XVIII (1988-1993) and was approved by the WTSC (Helsinki, March 1-12, 1993).

NOTES

1 As a consequence of a reform process within the International Telecommunication Union (ITU), the CCITT ceased to exist as of 28 February 1993. In its place, the ITU Telecommunication Standardization Sector (ITU-T) was created as of 1 March 1993. Similarly, in this reform process, the CCIR and the IFRB have been replaced by the Radiocommunication Sector.

In order not to delay publication of this Recommendation, no change has been made in the text to references containing the acronyms "CCITT, CCIR or IFRB" or their associated entities such as Plenary Assembly, Secretariat, etc. Future editions of this Recommendation will contain the proper terminology related to the new ITU structure.

2 In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

CONTENTS

GENERAL ASPECTS OF QUALITY OF SERVICE AND NETWORK PERFORMANCE IN DIGITAL NETWORKS, INCLUDING ISDNS

(Melbourne, 1988; revised Helsinki, 1993)

1 General

1.1 Purpose of Recommendation

This Recommendation has been developed to:

- provide descriptions of Quality of Service and network performance;
- illustrate how the Quality of Service and the network performance concepts are applied in digital networks, including ISDNs;
- describe the features of, and the relationship between these concepts;
- indicate and classify performance concerns for which parameters may be needed;
- identify generic performance parameters.

The generic term "performance" refers to Quality of Service and network performance as they are defined in 1.2.

1.2 Description of Quality of Service (QOS) and Network Performance (NP)

1.2.1 Description of Quality of Service

QOS is defined in Recommendation E.800 as follows: "Collective effect of service performances which determine the degree of satisfaction of a user of the service".

The note of Recommendation E.800 underlines that the QOS is characterized by to the combined aspects of:

- service support and service operability performance; and
- servability and service integrity performance.

The definition of Quality of Service in Recommendation E.800 is a wide one encompassing many areas of work, including subjective customer satisfaction. However, within this Recommendation the aspects of Quality of Service that are covered are restricted to the identification of parameters that can be directly observed and measured at the point at which the service is accessed by the user. Other types of QOS parameters which are subjective in nature, i.e. depend upon user actions or subjective opinions, will not be specified in the I-Series Recommendations on QOS.

1.2.2 Description of network performance

Network performance is measured in terms of parameters which are meaningful to the network provider and are used for the purposes of system design, configuration, operation and maintenance. NP is defined independently of terminal performance and user actions.

2 Purpose of QOS and NP

2.1 General

Bearer services and teleservices as described in the I.200-Series Recommendations are the objects which network and service providers offer to their customers. A major attribute of these services is the set of QOS parameters which a particular service offers. These parameters are user-oriented and take into account the elements involved in a particular service as given in Figure 2/I.211.

Bearer services and teleservices are supported by a range of connection types, each of which comprises several connection elements. The performance of the connection types is characterized by a set of NP parameters. These parameters are network oriented.

Figure 1 illustrates how the concepts of QOS and NP are applied in the ISDN environment.

2.2 Purpose of QOS

A typical user is not concerned with how a particular service is provided, or with any of the aspects of the network's internal design. However, he is interested in comparing one service with another in terms of certain universal, user-oriented performance concerns which apply to any end-to-end service. Therefore from a user's point of view, Quality of Service is best expressed by parameters which:

- focus on user-perceivable effects, rather than their causes within the network;
- do not depend, in their definition, on assumptions about the network internal design;
- take into account all aspects of the service from the user's point of view which can be objectively measured at the service access point;
- may be assured to a user at the service access point by the service provider(s);
- are described in network independent terms and create a common language understandable by both the user and the service provider.

As networks evolve, providers should maintain the Quality of Service of bearer services provided to users within acceptable limits, and where possible, should make Quality of Service improvements. The Quality of Service perceived by users should not be perceptibly degraded by network evolution.

2.3 Purpose of NP

A network provider is concerned with the efficiency and effectiveness of the network, in providing services to customers. Therefore from a network provider's point of view, NP is best expressed by parameters which provide information for:

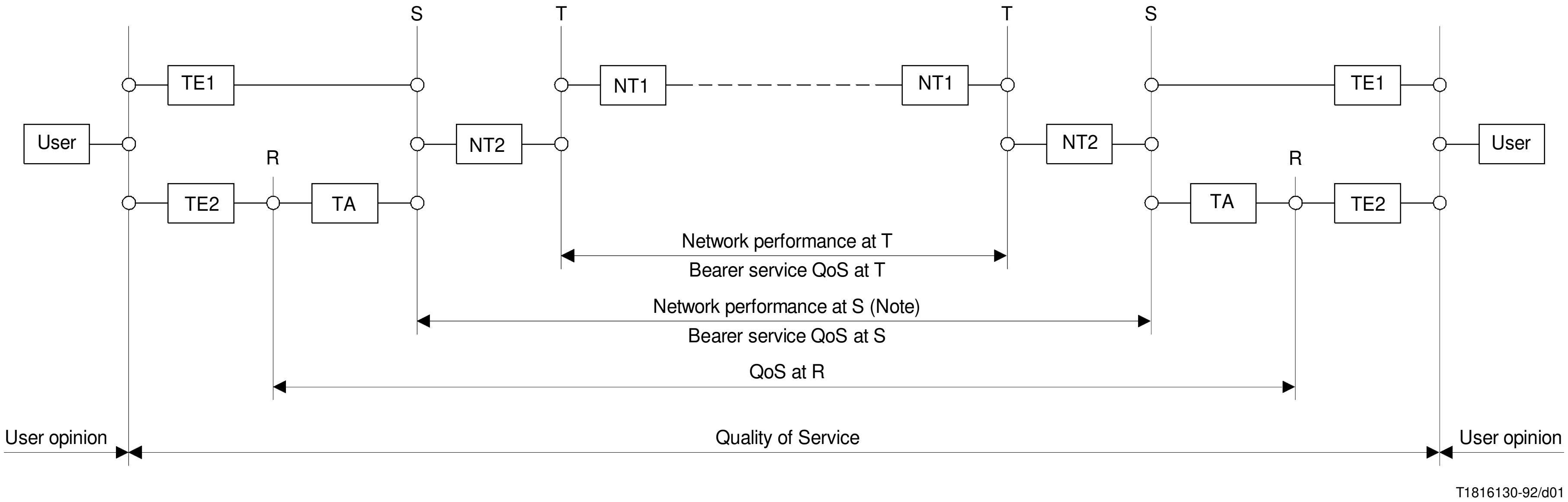
- system development;
- network planning, both nationally and internationally;
- operation and maintenance.

3 Principles for the development of QOS and NP parameters and values

3.1 General principles

3.1.1 Distinction between QOS and NP

The user-oriented QOS parameters provide a valuable framework for network design, but they are not necessarily usable in specifying performance requirements for particular connections. Similarly, the NP parameters ultimately determine the (user observed) QOS, but they do not necessarily describe that quality in a way that is meaningful to users. Both types of parameters are needed, and their values must be quantitatively related if a network is to be effective in serving its users. The definition of QOS and NP parameters should make mapping of values clear in cases where there is not a simple one-to-one relationship between them.



NOTE – Network performance at reference point S only appears when the S and T reference point is coincident.

FIGURE 1/I.350

General reference configuration for QoS and NP

Table 1 shows some of the characteristics which distinguish QOS and NP.

TABLE 1/I.350

Distinction between quality of service and network performance

Quality of service	Network performance
User oriented	Provider oriented
Service attribute	Connection element attribute
Focus on user-observable effects	Focus on planning, development (design), operations and maintenance
Between (at) service access points	End-to-end or network connection elements capabilities

3.1.2 Measurability of QOS and NP parameter values

Due to separating QOS and NP, a number of general points should be noted when considering the development of parameters:

- the definition of QOS parameters should be clearly based on events and states observable at service access points and independent of the network processes and events which support the service;
- the definition of NP parameters should be clearly based on events and states observable at connection element boundaries, e.g. protocol specific interface signals;
- the use of events and states in the definition of parameters should provide for measurements at the boundaries identified above. Such measurements should be verifiable in accordance with generally accepted statistical techniques.

3.1.3 Multiple network provider environments

It should be recognized in the development of parameter values that services may be provided by multiple providers. In such an environment different levels of QOS may be supported. Therefore, in practice, users may experience a variety of ranges of QOS. It is thus important to establish minimum performance levels for each service and for connection elements providing international connections.

3.2 QOS principles

For the definition of parameters for QOS in the ISDN, the concept of bearer services and teleservices needs to be borne in mind. In particular, there is a difference between the kinds of parameters which would describe the QOS of a bearer service and that of a teleservice, since the point of observation of, or access to, the service is different in each case. Figure 1 illustrates this point.

In the case of teleservices the interface between the user and the service provider may be a man-machine interface. In the case of bearer services this interface corresponds to the S or T reference points. As a result, some of the parameters for describing the QOS of a teleservice will be different from those which describe the QOS of a bearer service.

In describing the QOS of teleservicess, the performance of the terminal equipment (TE) has to be taken into account. For a teleservice, there should be a mapping between the QOS of the teleservice and the performance of the customer equipment including the terminal and the overall (end-to-end) NP of the connection elements supporting this service.

For a bearer service there should be mapping between the QOS of the bearer service and the overall (end-to-end) NP of connection elements supporting this service.

3.3 NP principles

When developing NP parameters the following points should be borne in mind:

- NP parameters must be measurable at the boundary of the network connection element(s) to which they are applied. The definitions should not be based on assumptions about either the internal characteristics of a network (or portions thereof), or the internal causes of impairments observed at the boundaries;
- the division of a network portion into sub-components should only be done if they must be specified separately in order to ensure satisfactory end-to-end performance or, where appropriate, to derive fair and reasonable allocations among providers. No network provider should bear a disproportionate cost in establishing and operating a service.

3.4 Primary and derived performance parameters

3.4.1 Description

- *Primary performance parameter*

  A parameter or a measure of a parameter determined on the basis of direct observations of events at services access points or connection element boundaries.

- *Derived performance parameter*

  A parameter or a measure of a parameter determined on the basis of observed values of one or more relevant primary performance parameters and decision thresholds for each relevant primary performance parameter.

3.4.2 Relationship between primary and derived performance parameters

A number of event types can be directly observed at service access points or connection element boundaries. Examples of such events are:

- the layer 3 protocol state transition associated with the transfer of a SETUP message or a DISCONNECT message across a connection element boundary;
- the correct receipt of an information bit (or a specified number of information bits) at an interface.

Parameters related to the time interval between specific events and the frequency of events can be measured. These directly measurable parameters or primary performance parameters describe the QOS (at service access points) or the NP (at connection element boundaries) during periods when the service or connection is available.

Derived performance parameters describe performance based on events which are defined as occurring when the value of a function of a primary performance parameter(s) crosses a particular threshold. These derived threshold events identify the transitions between the available and the unavailable states. Parameters related to the time interval between these derived threshold events and their frequency can be identified. These derived performance parameters describe the QOS and the NP for all time intervals; i.e. during periods when the service or connection is available or unavailable.

NOTE – Primary performance parameters are measured for all time intervals, since the transitions between available and unavailable states depend upon the value of these parameters. However, the values of primary performance parameters would not be specified for a service or connection in the unavailable state.

4 Generic performance parameters

Nine generic primary performance parameters are listed below. These have been developed as a result of the matrix approach described in Annex A. These parameters may be used in developing specific QOS and NP parameters:

- access speed;
- access accuracy;
- access dependability;
- information transfer speed;

– information transfer accuracy;

– information transfer dependability;

– disengagement speed;

– disengagement accuracy;

– disengagement dependability.

Subclause 3.4 defines derived performance parameters in addition to primary parameters. Derived performance parameters are determined utilizing a function of the primary performance parameter values. Recommendation G.821 defines one such function, which identifies transitions between available and unavailable states based on a threshold for severely errored seconds. The generic derived performance parameter associated with such a function is availability.

Examples of specific primary and derived performance parameters for bearer service QOS and those for circuit-switched and packet-switched NP are provided in Annex B.

Annex A

Method of identifying parameters

(This annex forms an integral part of this Recommendation)

A.1 The matrix approach

The matrix provides a systematic method of identifying and organizing candidate network performance parameters with the objective of defining a concise set of parameters and, where appropriate, their QOS counterparts. This tool should be used as the basis for collection and evaluation of network performance parameters for digital networks, including ISDNs.

A.2 3 × 3 matrix approach for network performance

The 3 × 3 matrix approach for network performance is illustrated in Figure A.1. The main features are as follows:

1) Each row represents one of the three basic and distinct communication functions.

   NOTE – The access function represents the connectionless as well as connection-oriented services which are possible with ISDNs.

2) Each column represents one of the three mutually exclusive outcomes possible when a function is attempted.

3) The 3 × 3 matrix parameters are defined on the basis of events at connection element boundaries and are termed "primary performance parameters". "Derived performance parameters" are defined on the basis of a functional relationship of primary performance parameters, outage thresholds and an observation interval.

4) NP primary performance parameters should be defined so as to be measurable at the boundaries of the connection element(s) to which they apply. NP parameter definitions should not depend upon assumptions about impairment causes that are not detectable at the boundaries.

5) Availability is a derived performance parameter. Decisions on the appropriate primary performance parameters, outage threshold and algorithms for its definition require further detailed study.

   NOTE – The following terminology problems are pointed out. Appropriate terms should be selected after further study:

a) The term "access" is used. However, the term "selection (of the connection type, the destination and facility)" has been proposed as an alternative.

b) The term “dependability” is used. However, the definition of dependability as used here is somewhat different from that in Recommendation G.106 *(Red Book).* Alternative terms, “inserveability” and “refusal” are proposed.

c) The term “availability” is provisionally used. An alternative term “acceptability” has been proposed.

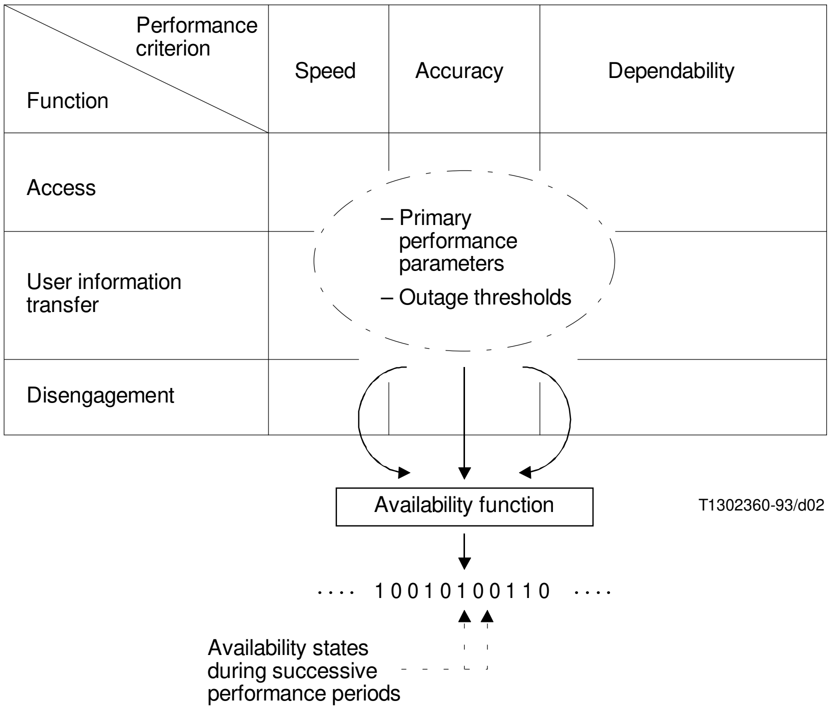


FIGURE A.1/I.350

3 × 3 matrix approach and determination of availability states

A.3 3 × 3 matrix approach for QOS

The same 3 × 3 matrix approach as that described for network performance may be used for the related Quality of Service parameters.

QOS parameters should be defined so as to be measurable at service access points. QOS parameter definitions should not depend upon assumptions of impairment causes that are not detectable at the service access points.

Loss of service parameters are considered to be derived QOS parameters. An alternative matrix has been proposed and is still under consideration.

A.4 Description of the basic communication functions

A.4.1 Access

The access function begins upon issuance of an access request signal or its implied equivalent at the interface between a user and the communication network. It ends when either:

1) a ready for data or equivalent signal is issued to the calling users; or

2) at least one bit of user information is input to the network (after connection establishment in connection-oriented services).

It includes all activities traditionally associated with physical circuit establishment (e.g. dialling, switching, and ringing) as well as any activities performed at higher protocol layers.

A.4.2 User information transfer

The user information transfer begins on completion of the access function, and ends when the "disengagement request" terminating a communication session is issued. It includes all formatting, transmission, storage, error control and media conversion operations performed on the user information during this period, including necessary retransmission within the network.

A.4.3 Disengagement

There is a disengagement function associated with each participant in a communication session: each disengagement function begins on issuance of a disengagement request signal. The disengagement function ends, for each user, when the network resources dedicated to that user's participation in the communication session have been released. Disengagement includes both physical circuit disconnection (when required) and higher-level protocol termination activities.

A.5 Description of the performance

A.5.1 Speed

Speed is the performance criterion that describes the time interval that is used to perform the function or the rate at which the function is performed. (The function may or may not be performed with the desired accuracy.)

A.5.2 Accuracy

Accuracy is the performance criterion that describes the degree of correctness with which the function is performed. (The function may or may not be performed with the desired speed.)

A.5.3 Dependability

Dependability is the performance criterion that describes the degree of certainty (or surety) with which the function is performed regardless of speed or accuracy, but within a given observation interval.

Annex B

Relationship between generic and possible specific QOS and NP parameters

(This annex forms an integral part of this Recommendation)

This annex illustrates the qualitative relationship between the generic parameters defined in this Recommendation and a candidate set of specific QOS and NP parameters. Tables B.1, B.2, B.3 and B.4 illustrate the relationship between the generic parameters and specific bearer service QOS, circuit-switched NP, and packet-switched NP parameters, and B-ISDN NP parameters, respectively.

In general, only those network performance parameters whose values can be significantly altered by the allocation of network resources to a specified ATM connection are reasonable candidates for adjustment in support of a negotiated QOS. The following ATM layer network performance parameters may be adjusted on a unidirectional basis for a specified ATM connection in support of a negotiated QOS:

1) cell loss ratio;

2) cell transfer capacity.

   a) first value of cell transfer capacity;

   b) second value of cell transfer capacity.

By use of the cell loss priority bit in the header of an ATM cell, it seems possible to choose between two values for the performance objective for the cell loss ratio measured on that ATM connection. This assumes that the cell loss ratio is significantly affected by congestion.

A single value of cell transfer capacity appears sufficient to characterize the throughput capacity of an ATM connection that carries constant bit rate (CBR) traffic.

For variable bit rate (VBR) traffic, it appears useful to also permit the negotiation of a second value of cell transfer capacity. The use of two values can permit the network to derive greater efficiency from the statistical multiplexing of VBR traffic. The exact interpretation of these two values is not resolved at this time.

Other ATM performance parameters pertaining to cell transfer have been defined, but are more difficult to change in support of a negotiated QOS for an ATM connection. Parameters in this category include the cell transfer delay, cell delay variation, cell error ratio and cell misinsertion rate.

TABLE B.1/I.350

Qualitative relationship between generic performance parameters and candidate bearer service Q

Generic parameters		Bearer service QOS parameters									
		Primary performance parameters									
		Access delay	Incorrect access probability	Access denaial probability	User informa-tion transfer rate	User informa-tion transfer rate	User informa-tion error probability	Extra user informa-tion delivery probability	User informa-tion misde-livery probability	User informa-tion loss probability	Dis gage de
Primary	Access speed	X									
	Access accuracy		X								
	Access dependability			X							
	Information transfer speed				X	X					
	Information transfer accuracy						X	X	X		
	Information transfer dependabilty									X	
	Disengagement speed										X
	Disengagement accuracy										
	Disengagement dependability										
Derived	Availability										

TABLE B.2/I.350

Qualitative relationship between generic performance parameters and candidate circuit-switched

Generic parameters		Circuit-switched NP parameters – Primary performance parameters: Connection set-up delay	Alerting delay	Connection set-up error probability	Connection set-up denial probability	Propagation delay	Degraded minutes	Severely errored seconds	Errored seconds	Disconnected delay
Primary	Access speed	X	X							
	Access accuracy			X						
	Access dependability				X					
	Information transfer speed					X				
	Information transfer accuracy						X	X	X	
	Information transfer dependabilty									
	Disengagement speed									X
	Disengagement accuracy									
	Disengagement dependability									
Derived	Availability									

TABLE B.3/I.350

Qualitative relationship between generic performance parameters and candidate packet-switched

Generic parameters		Packet-switched NP parameters									
		Primary performance parameters									
		Virtual circuit set-up delay	Vitrual circuit set-up error probability	Vitrual circuit set-up denial probability	Data packet transfer delay	Through-put capacity	Residual error rate	Reset probability	Reset stimulus probability	Virtual circuit clearing delay	p
Primary	Access speed	X									
	Access accuracy		X								
	Access dependability			X							
	Information transfer speed				X	X					
	Information transfer accuracy						X	X	X		
	Information transfer dependability						X	X	X		
	Disengagement speed									X	
	Disengagement accuracy										
	Disengagement dependability										
Derived	Availability										

TABLE B.4/I.350

Qualitative relationship between generic performance parameters and candidate B-ISDN NP

Generic parameters		B-ISDN NP parameters										
		Primary performance parameters										
		Connection set-up delay	Misrouted connection ratio	Connection set-up denial ratio	Cell transfer delay	Cell delay variation	Cell transfer capacity	Cell error ratio	Severely errored cell block ratio	Cell loss ratio	Cell mis-insertion rate	Di
Primary	Access speed	X										
	Access accuracy		X									
	Access dependability			X								
	Information transfert speed				X	X	X					
	Information transfer accuracy							X	X		X	
	Information transfer dependability									X		
	Disengagement speed											
	Disengagement accuracy											
	Disengagement dependability											
Derived	Availability											