INTERNATIONAL TELECOMMUNICATION UNION

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

X.641

(12/97)

SERIES X: DATA NETWORKS AND OPEN SYSTEM COMMUNICATION

OSI networking and system aspects – Quality of Service

Information technology – Quality of Service: Framework

ITU-T Recommendation X.641

(Previously CCITT Recommendation)

ITU-T X-SERIES RECOMMENDATIONS

DATA NETWORKS AND OPEN SYSTEM COMMUNICATION

PUBLIC DATA NETWORKS	
Services and facilities	X.1–X.19
Interfaces	X.20–X.49
Transmission, signalling and switching	X.50–X.89
Network aspects	X.90–X.149
Maintenance	X.150–X.179
Administrative arrangements	X.180–X.199
OPEN SYSTEM INTERCONNECTION	
Model and notation	X.200–X.209
Service definitions	X.210–X.219
Connection-mode protocol specifications	X.220–X.229
Connectionless-mode protocol specifications	X.230–X.239
PICS proformas	X.240–X.259
Protocol Identification	X.260–X.269
Security Protocols	X.270–X.279
Layer Managed Objects	X.280–X.289
Conformance testing	X.290–X.299
INTERWORKING BETWEEN NETWORKS	
General	X.300–X.349
Satellite data transmission systems	X.350–X.399
MESSAGE HANDLING SYSTEMS	X.400–X.499
DIRECTORY	X.500–X.599
OSI NETWORKING AND SYSTEM ASPECTS	
Networking	X.600–X.629
Efficiency	X.630–X.639
Quality of service	**X.640–X.649**
Naming, Addressing and Registration	X.650–X.679
Abstract Syntax Notation One (ASN.1)	X.680–X.699
OSI MANAGEMENT	
Systems Management framework and architecture	X.700–X.709
Management Communication Service and Protocol	X.710–X.719
Structure of Management Information	X.720–X.729
Management functions and ODMA functions	X.730–X.799
SECURITY	X.800–X.849
OSI APPLICATIONS	
Commitment, Concurrency and Recovery	X.850–X.859
Transaction processing	X.860–X.879
Remote operations	X.880–X.899
OPEN DISTRIBUTED PROCESSING	X.900–X.999

For further details, please refer to ITU-T List of Recommendations.

INTERNATIONAL STANDARD 13236

ITU-T RECOMMENDATION X.641

INFORMATION TECHNOLOGY – QUALITY OF SERVICE: FRAMEWORK

Summary

This Recommendation | International Standard provides a common basis for the coordinated development and enhancement of a wide range of standards that specify or reference Quality of Service (QOS) requirements or mechanisms. It offers a means of developing or enhancing standards related to QOS and provides concepts and terminology that will assist in maintaining the consistency of related standards.

Source

The ITU-T Recommendation X.641 was approved on the 12th of December 1997. The identical text is also published as ISO/IEC International Standard 13236.

FOREWORD

ITU (International Telecommunication Union) is the United Nations Specialized Agency in the field of telecommunications. The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of the ITU. 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, establishes the topics for study by the ITU-T Study Groups which, in their turn, produce Recommendations on these topics.

The approval of Recommendations by the Members of the ITU-T is covered by the procedure laid down in WTSC Resolution No. 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

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

INTELLECTUAL PROPERTY RIGHTS

The ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. The ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, the ITU had not received notice of intellectual property, protected by patents, which may be required to implement this Recommendation. However, implementors are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database.

CONTENTS

Introduction

The purpose of this Recommendation | International Standard is to provide a common basis for the coordinated development and enhancement of the wide range of standards that specify or reference Quality of Service (QOS) requirements or mechanisms in an Information Technology (IT) environment. It offers a means of developing or enhancing standards relating to QOS and provides concepts and terminology that will assist in maintaining the consistency of related standards. This Recommendation | International Standard is complementary to existing ITU-T Recommendations which define performance objectives and network signalling of QOS and it is not the intention of this Framework to lead to a need to revise any such Recommendations.

The initial work in developing this Framework for QOS was done with the objective of supplementing and clarifying the description of QOS contained in the Basic Reference Model of Open Systems Interconnection (OSI) (see ITU-T Rec. X.200 | ISO/IEC 7498-1). It is recognised, however, that management of QOS is important not only in OSI communications but also in a much wider context, and that there is value in encouraging a common approach to QOS that can extend to other IT and communications architectures, to distributed processing in general and to Open Distributed Processing (ODP) in particular.

Hence this Recommendation | International Standard is structured and written in such a way as to make it easy for many communities to adopt its approach, concepts, terminology and definitions. Its concepts and terms are defined without reference to any particular architecture, so that they can be adopted and applied by other communities to a variety of architectures and protocols. This general treatment is supplemented by examples from OSI, ODP and elsewhere.

To assist the OSI community, Annex A defines how the general framework applies to the specific case of OSI communications, both peer-to-peer and multi-peer.

This QOS Framework contains an introduction, a scope and field of application and a set of QOS-related definitions and abbreviations. The concepts of QOS are introduced in clause 5, which also highlights user requirements. Clause 6 defines QOS characteristics with respect to the user requirements. Clauses 7, 8 and 9 cover QOS management, QOS mechanisms and the expression of specific QOS requirements, respectively. QOS verification is discussed in clause 10, and conformance, consistency and compliance are discussed in clause 11.

Annexes are provided which stipulate:

- the model of QOS for OSI;
- statistical derivations of characteristics;
- a standards structure with respect to QOS;
- a discussion of the issue of ‘cost’; and
- a bibliography.

Other standards communities are encouraged to study this Recommendation | International Standard and, in the interests of consistency, to consider the adoption of the parts that are relevant to their field, when this can be accomplished to good effect, without destabilisation of existing Recommendations | International Standards.

INTERNATIONAL STANDARD

ITU-T RECOMMENDATION

INFORMATION TECHNOLOGY – QUALITY OF SERVICE: FRAMEWORK

1 Scope

This QOS Framework is a structured collection of concepts and their relationships which describes QOS (Quality of Service) and enables the partitioning of, and relationships between, the topics relevant to QOS in Information Technology (IT) to be expressed by a common means of description. In particular, this QOS Framework is directed at IT systems and their use in providing Open Distributed Processing services.

This QOS Framework is intended to assist those designing and specifying IT systems, and those defining communications services and protocols, by providing guidance on QOS applicable to systems, services and resources of various kinds. It describes how QOS can be characterized, how QOS requirements can be specified, and how QOS can be managed.

This QOS Framework defines terminology and concepts for QOS in IT. It introduces the concept of QOS characteristics, which represent the fundamental aspects of QOS that are to be managed in various ways; and it defines a number of QOS characteristics of particular importance. These definitions are independent of how QOS is represented or controlled in a real system.

This Framework describes how QOS requirements can be expressed, and identifies a number of QOS mechanisms (such as three-party negotiation) that can be used as components of QOS management functions to meet QOS requirements of various kinds. It also describes the circumstances in which various combinations of mechanisms may be appropriate.

This QOS Framework provides a basis for the specification of extensions and enhancements to existing or planned standards, as a result of the need for, and application of, the QOS concepts defined in this Recommendation | International Standard. It is not the intent of this Recommendation | International Standard to destabilise any existing Recommendations | International Standards; rather, it is intended that this QOS Framework may be used by:

- developers of new or revised IT-related standards which define or use QOS mechanisms; and
- IT users expressing requirements for QOS.

This QOS Framework does not attempt to provide a basis for the specification of performance objectives or network signalling of QOS in public communications networks. The QOS aspects of these communications services are addressed by other ITU-T Recommendations.

The intent of this Recommendation | International Standard is to provide a common vocabulary to both service providers and service users. Nothing in this Recommendation | International Standard should be construed as placing requirements on either service providers or service users. It is hoped that a common approach and vocabulary for QOS will assist multiple service providers to deliver end-to-end QOS to end-systems.

This QOS Framework specifically excludes the detailed specification of QOS mechanisms. It is not the intent of this Recommendation | International Standard to serve as an implementation specification, to be a basis for appraising the conformance of implementations, or to define particular services and protocols. Rather, it provides a conceptual and functional framework for QOS which allows independent teams of experts to work productively on the development of Recommendations | International Standards.

As applied to OSI, this QOS Framework is consistent with the OSI Basic Reference Model in that it describes operations and mechanisms which are assignable to layers as specified in the OSI Basic Reference Model. It is consistent with the OSI Management Framework (see ITU-T Rec. X.700 | ISO/IEC 7498-4) and the Systems Management Overview (see ITU-T Rec. X.701 | ISO/IEC 10040) in its assignment of functions to management entities. In Annex A, this QOS Framework presents a model of QOS for OSI which identifies the entities that participate in the management of QOS, defines the flow of QOS-related information between them and describes how this information is used.

2 Normative references

The following Recommendations and International Standards contain provisions which, through reference in this text, constitute provisions of this Recommendation | International Standard. At the time of publication, the editions indicated were valid. All Recommendations and Standards are subject to revision, and parties to agreements based on this Recommendation | International Standard are encouraged to investigate the possibility of applying the most recent edition of the Recommendations and Standards listed below. Members of IEC and ISO maintain registers of currently valid International Standards. The Telecommunication Standardization Bureau of the ITU maintains a list of currently valid ITU-T Recommendations.

2.1 Identical Recommendations | International Standards

- ITU-T Recommendation X.200 (1994) | ISO/IEC 7498-1:1994, *Information technology – Open Systems Interconnection – Basic Reference Model: The Basic Model.*
- ITU-T Recommendation X.210 (1993) | ISO/IEC 10731:1994, *Information technology – Open Systems Interconnection – Basic Reference Model: Conventions for the definition of OSI services.*
- ITU-T Recommendation X.746 (1995) | ISO/IEC 10164-15:1995, *Information technology – Open Systems Interconnection – Systems management: Scheduling function.*
- ITU-T Recommendation X.902 (1995) | ISO/IEC 10746-2:1996, *Information technology – Open distributed processing – Reference Model: Foundations.*

2.2 Paired Recommendations | International Standards equivalent in technical content

- CCITT Recommendation X.700 (1992*), Management framework for Open Systems Interconnection (OSI) for CCITT applications.*

  ISO/IEC 7498-4: 1989, *Information processing systems – Open Systems Interconnection – Basic Reference Model – Part 4: Management Framework*.

2.3 Additional references

- CCITT Recommendation X.140 (1992), *General Quality of Service parameters for communication via public data networks*.

3 Definitions

For the purposes of this Recommendation | International Standard, the following definitions apply.

3.1 Basic Reference Model of Open Distributed Processing definitions

This Recommendation | International Standard uses the following term drawn from ITU-T Rec. X.902 | ISO/IEC 10746-2:

- Quality of Service (QOS): A set of qualities related to the collective behaviour of one or more objects.

3.2 Basic Reference Model – OSI Management Framework definitions

This Recommendation | International Standard uses the following term drawn from CCITT Rec. X.700 | ISO/IEC 7498-4:

- Managed Object (MO): The OSI Management view of a resource within the OSI Environment that may be managed through the use of OSI Management protocol(s).

3.3 Service Conventions definitions

This Recommendation | International Standard uses the following terms drawn from ITU-T Rec. X.210 | ISO/IEC 10731:

- (N)-service-facility;
- (N)-service-provider;
- (N)-service-user;

– Requester;

– Acceptor.

NOTE – In this Recommendation | International Standard, *service provider* is a generic term to indicate the provision or use of a service at some point in a system. It should not be confused with the provision of a commercial service by any commercial agency.

3.4 OSI Basic Reference Model definitions

This Recommendation | International Standard uses the following terms drawn from ITU-T Rec. X.200 | ISO/IEC 7498-1:

– (N)-entity;

– (N)-layer;

– (N)-protocol;

– (N)-protocol-data-unit;

– (N)-service;

– (N)-service-access-point;

– (N)-subsystem;

– open system;

– OSI environment.

3.5 QOS Framework definitions

3.5.1 QOS concepts and modelling definitions

3.5.1.1 QOS category: A group of user requirements that leads to the selection of a set of QOS requirements.

3.5.1.2 QOS characteristic: A quantifiable aspect of QOS, which is defined independently of the means by which it is represented or controlled.

3.5.1.3 QOS management: Any set of activities performed by a system or communications service to support QOS monitoring, control and administration.

3.5.1.4 QOS mechanism: A specific mechanism that may use protocol elements, QOS parameters or QOS context, possibly in conjunction with other QOS mechanisms, in order to support establishment, monitoring, maintenance, control, or enquiry of QOS.

3.5.1.5 QOS of OSI communications: A set of qualities related to the provision of an (N)-service, as perceived by an (N)-service-user.[1)]

3.5.1.6 QOS policy: A set of rules that determine the QOS characteristics and QOS management functions to be used.

3.5.2 Information-oriented definitions

3.5.2.1 QOS context: QOS information that is retained, interpolated or extrapolated by one or more entities and used in managing QOS: it is further classified into requirement context and data context.

3.5.2.2 QOS data: QOS information other than QOS requirements, e.g. warnings, QOS measures and information used in QOS enquiries.

3.5.2.3 QOS information: Information related to QOS: it is classified into QOS context (when retained in an entity) and QOS parameters (when conveyed between entities); and it is classified into QOS requirements (if it expresses a requirement for QOS) and QOS data (if it does not).

3.5.2.4 QOS measure: One or more observed values relating to a QOS characteristic.

3.5.2.5 QOS parameter: QOS information that is conveyed between entities as part of a QOS mechanism; parameters are classified into requirement parameters and data parameters; the information conveyed may relate to one or more QOS characteristics.

---

1) This definition of QOS is a specialisation of the definition of QOS given in 3.1 above, as applied to OSI communications.

3.5.2.6 QOS requirement: QOS information that expresses part or all of a requirement to manage one or more QOS characteristics, e.g. a maximum value, a target, or a threshold; when conveyed between entities, a QOS requirement is expressed in terms of QOS parameters.

3.5.2.7 QOS operating target: QOS information that represents the target values of a set of QOS characteristics, derived from QOS requirements.

3.5.3 Management function definitions

3.5.3.1 QOS alert: The use of QOS mechanisms to signal to an entity that some limit has been reached or threshold crossed.

3.5.3.2 QOS attribute: An attribute of a managed object relating to QOS.

3.5.3.3 QOS control: The use of QOS mechanisms to modify conditions so that a desired set of QOS characteristics is attained for some systems activity, while that activity is in progress.

3.5.3.4 QOS enquiry: The use of QOS mechanisms to determine properties of the environment relating to QOS.

3.5.3.5 QOS establishment: The use of QOS mechanisms to create the conditions for some systems activity, before that activity occurs, so that a desired set of QOS characteristics is attained.

3.5.3.6 QOS maintenance: The use of QOS mechanisms to maintain a set of QOS characteristics at required values for some systems activity, while the activity is in progress.

3.5.3.7 QOS management function: A function specifically intended to meet a user or application requirement for QOS, provided by one or more QOS mechanisms.

3.5.3.8 QOS monitoring: The use of QOS measures to estimate the values of a set of QOS characteristics actually achieved for some systems activity.

4 Abbreviations

For the purposes of this Recommendation | International Standard, the following abbreviations apply:

ASO	Application Service Object
CHQ	Controlled Highest Quality
CL	Connectionless
CO	Connection-oriented
GDMO	Guidelines for the Definition of Managed Objects
IT	Information Technology
LAN	Local Area Network
LQA	Lowest Quality Acceptable
MTBF	Mean Time Between Failures
MTTR	Mean Time to Repair
(N)-PCF	(N)-Policy Control Function
(N)-PDU	(N)-protocol-data-unit
(N)-PE	(N)-protocol-entity
(N)-QCF	(N)-QOS Control Function
(N)-SAP	(N)-service-access-point
(N)-SDU	(N)-service-data-unit
ODP	Open Distributed Processing
OSI	Open Systems Interconnection
OSIE	OSI Environment
PDU	Protocol Data Unit

QMF	QOS Management Function
QOS	Quality of Service
SAP	Service Access Point
SDU	Service Data Unit
SNPA	Sub-network Point of Attachment
SP	System Performance
SQCF	System QOS Control Function
TCCA	Time-Critical Communications Architecture
TCCS	Time-Critical Communications System
TPDU	Transport PDU
TSAP	Transport SAP
UIU	User Information Unit

5 Concepts of this QOS Framework

5.1 Introduction

This clause develops a set of fundamental concepts for QOS. It contains a description of concepts for the information and functional aspects of QOS including:

- the concept of service to which QOS applies;
- QOS characteristics that describe the fundamental aspects of QOS that are to be managed;
- QOS requirements and policies and the QOS management functions that realize them;
- basic QOS mechanisms that combine to form QOS management functions;
- QOS categories that represent the particular sets of user requirements for QOS imposed by certain environments (such as time-critical communications), or by systems-level policy;
- the stages of an activity at which QOS can be managed.

These concepts are described in a way that highlights the operation of QOS management. It begins by discussing the inherent QOS characteristics and then considers how QOS requirements drive the selection and use of QOS management functions and QOS mechanisms.

Figure 5-1 shows the relationships between the fundamental QOS concepts.

5.2 The service to which QOS applies

In this QOS Framework, as in the expression 'Quality of Service', the term 'service' is to be understood in a very general sense so as to permit the widest possible application of the Framework. Specifically, it includes (but is not necessarily limited to):

- the provision of processing and information repository functions by entities, objects, applications, applications processes, etc.; for example, time-delay and reliability-related characteristics apply to these;

  NOTE – the terms used to denote these entities, etc., depend upon the architectural models applicable to the particular distributed systems environments in which QOS is to be represented: the term 'entity' is used in the Framework in a neutral sense, i.e. with no implication for the choice of architectural model.

- interactions between entities, objects, applications, etc.;
- information held in the system; for example, confidentiality and lifetime characteristics apply to information;
- communications services;
- (the possibility of use of) physical equipment.

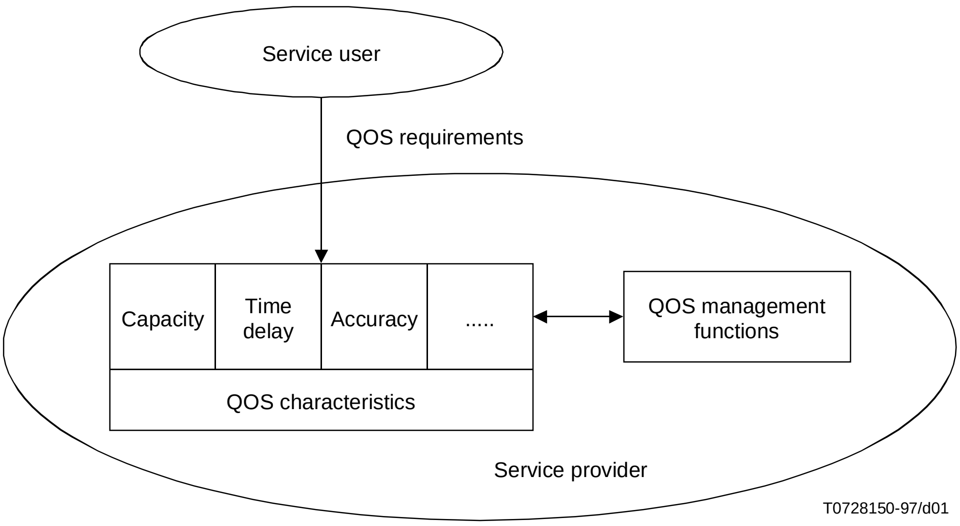


Figure 5-1 – Relationships between QOS concepts

5.3 QOS characteristics

The term QOS characteristic is used in this QOS Framework in a particular sense. A QOS characteristic represents some aspect of the QOS of a system, service or resource that can be identified and quantified. It denotes the true underlying state of affairs for the item, as opposed to any measurement or control parameter. QOS characteristics are defined independently of the means by which they are represented or controlled.

QOS characteristics are intended to be used to model the actual, rather than the observed, behaviour of the systems that they characterize.

NOTE – For example, the characteristic of transit delay (of something) between two points is the actual time that occurs between the instants of passage of that thing past those points. This transit delay can never be known exactly, although it can be approximated by measurement. Further, requirements can be stated concerning transit delay – such as that it must not exceed a specified value.

Some QOS characteristics are considered to be generic characteristics, some are specialisations of these characteristics and others are derived characteristics. (For further discussion of QOS characteristics see clause 6.)

5.4 User QOS requirements and QOS policies

QOS management activities are driven by user requirements, the systems and communications environment, and the policies that are in force for the activity. User requirements are quantified and expressed as a set of QOS requirements (which is a type of QOS information).

Although a user's requirements may vary considerably in detail between individual instances of an activity, the QOS characteristics of interest and the QOS management functions employed will typically be determined by the type of system and distributed application that are to be supported and the types of network technology used. It is not expected that every system or network will implement every or any type of QOS mechanism. Systems and communications networks will be designed, procured or configured in accordance with one or more QOS policies, which will determine which QOS characteristics and QOS management functions are to be used.

Some systems and networks will need to be configurable to operate with different QOS policies, requiring that different sets of QOS management functions be available.

5.5 QOS requirements, QOS parameters and QOS context

A user requirement originates with a user entity that wishes to use a service, such as communications, and is formulated into one or more QOS requirements. These QOS requirements can be expressed as QOS parameters (when they need to be conveyed between entities) and QOS context (when they are retained in an entity).

In general, an activity is initiated by a user entity, whose requirements for QOS are either conveyed dynamically to the provider of the service as QOS parameters, or made available to the service provider as QOS context, or a combination of the two. The QOS parameters are conveyed to some or all of the entities involved in providing the service and possibly also to the corresponding application process.

Entities that receive QOS requirements analyse them and determine the QOS management functions or mechanisms that are required to meet them. This may involve generating further, typically more detailed, QOS requirements and conveying them to other entities as QOS parameters. The receiving entities analyse the QOS requirements they receive, and may generate yet further QOS requirements to be conveyed to other entities, and so on. One common example of this process is QOS negotiation across multiple layers during OSI connection establishment.

Thus, in general, a QOS parameter is a vector or scalar value that is conveyed between entities, either in the same system or in different systems.

NOTE – In this Recommendation | International Standard the use of the term 'QOS parameter' is specifically limited to conveyed values. Some documents do not distinguish between QOS characteristics and QOS parameters, and use the term QOS parameters for both.

For example, it may be a QOS requirement that the transit delay of data units between two points should not exceed a stated maximum, or that the average transit delay should be close to a stated target. In such cases, it is the true transit delay that is the characteristic. The QOS requirement will be expressed as QOS parameters (or QOS context) giving maximum or target values, for example. The interactions between entities will use QOS parameters to convey the relevant QOS information.

Depending on the exact requirement, the QOS parameters conveyed (or the QOS context retained) may be of different kinds, including:

- a desired level of characteristic, i.e. a target of some kind;
- a maximum or minimum level of a characteristic, i.e. a limit;
- a measured value, used to convey historical information;
- a threshold level;
- a warning or signal to take corrective action; or
- a request for operations on managed objects relating to QOS, or the results of such operations.

QOS requirements may relate to a number of QOS characteristics and, at least in principle, express trade-offs between them.

QOS requirements may apply to a single instance of information transfer or interaction, or may apply to multiple transfers or interactions (e.g. over a given period of time, for the duration of a connection or association, or for the duration of some longer provision of service, such as a customer subscription).

5.6 QOS management functions and QOS mechanisms

QOS management refers to all the activities relating to the control and administration of QOS within a system or network.

The term QOS management function (QMF) refers to any function designed to assist in satisfying one or more user QOS requirements. QMFs will in general have a number of components, which are called QOS mechanisms (e.g. three-party QOS negotiation).

A QOS mechanism is performed by one or more entities to meet one or more QOS requirements. Thus QOS mechanisms are driven by QOS requirements expressed as QOS parameters received by the entities that perform them, or made available to them as QOS context. The operation of a QOS mechanism may be local processing (e.g. reservation of resources, setting window sizes, etc.) which by itself meets the QOS requirement; or it may involve the generation of further QOS requirements and their communication to other entities, as described in 5.5. It may lead to the invocation of further QOS mechanisms.

The activities that may be supported by QMFs include:

- establishment of QOS for a set of QOS characteristics;
- monitoring of the observed values of QOS;
- maintenance of the actual QOS as close as possible to the target QOS;
- control of QOS targets;
- enquiry upon some QOS information or action; and
- alerts as a result of some event relating to QOS management.

NOTE – QOS requirements are commonly met (in current static environments) by being built into the systems configuration by systems design, by systems dimensioning, by the procurement of suitable services, or by dedicating resources. In more dynamic environments, increased use of QMFs is likely to be necessary to establish the environment, to monitor and control the environment during use and to capture historic information about system operation.

5.7 QOS categories

Different user application types will have different requirements for establishing QOS and for controlling and maintaining the actual QOS achieved. For example, the QOS requirements for video streams are typically very different from those for database update transactions.

These different types of user or application requirement, which are termed QOS categories, lead to the choice of particular sets of QOS characteristics to be managed. Subclause 6.4 identifies a number of fundamental QOS categories.

5.8 Initiation of QOS management

QOS management involves the use of different QMFs at different points in a system activity. The QOS requirements for a particular activity or set of activities can be expressed and/or conveyed in different ways and at different times in relation to the events or activities whose QOS is to be managed.

QOS management would be used at the following stages of an activity (related to 7.1.1):

- before initiation, when predictions are made about the QOS situation to gauge what mechanisms may be necessary to achieve some objective;
- at initiation of the activity – The QOS requirements can be negotiated between the service users and service provider at establishment time (e.g. when a connection is being established); and
- during the activity – The QOS requirements may change during the operational period of the activity due to changed requirements, detected performance loss, explicit indications from the service provider, or explicit indications from one or more third parties.

For any specific activity, the selection of the most appropriate stages to carry out QOS management depends on the type of QOS requirement and the duration of the activities to which it relates.

6 Definition of QOS characteristics

6.1 Introduction

This clause develops the concept of QOS characteristics (in 6.2) and defines a number of QOS characteristics of importance to communications and processing (in 6.3).

A QOS characteristic is a quantifiable aspect of QOS, which is defined independently of the means by which it is represented or controlled. It represents the true underlying state of affairs, as opposed to any measurement or control parameter, and can therefore be thought of as a quantity in a mathematical model of a (distributed) system. Thus, in defining a QOS characteristic such as throughput, the intention is to say what throughput means. This is distinct from how it may be measured, controlled, requested, negotiated, and so on, which are discussed in later subclauses.

The values that may be taken by QOS characteristics include not only numbers (e.g. Booleans, integers, reals, complex numbers, etc.), but also vectors, matrices, ranks and names of states.

Efforts are made to achieve maximum consistency of definition across different characteristics by defining ‘generic’ characteristics, and then both specialising them to particular environments and deriving others from them. This is discussed in 6.2.1 below.

It is recognised that the characteristics defined in this clause will not meet all future QOS requirements. Hence 6.2.2 gives guidance on how QOS characteristics should be defined. The definitions in 6.3 follow this guidance.

This clause also identifies a number of QOS categories.

6.2 Aspects of QOS characteristic definition

6.2.1 Generic, specialised and derived characteristics

6.2.1.1 Specialisation

Many QOS characteristics can be applied to a variety of circumstances. For example, one can define the transit delay of frames in a Local Area Network (LAN) supporting a real-time process-control environment, or of network access protocol PDUs between two SNPAs. Similarly one can define the throughput of a connection, or of any other communications channel.

In such cases it is important to have a common underlying definition of the characteristic that can be applied to all the particular circumstances. To achieve this, the first step is to define a 'generic characteristic' independently of what it is applied to; and the second is to define various 'specialisations' that may or must be applied in order to make the characteristic concrete and useable in practice. So, for example, 6.3 first defines *time delay* as a generic characteristic. It then defines some specialised characteristics from *time delay*, one of which is *transit delay*. It then identifies various further specialisations that may or must be applied to make the characteristic concrete; these include specifying the type of data transferred, the points between which transit is defined, and so on.

Thus, there can be several levels of specialisation of a characteristic, for example:

- time delay;
- transit delay;
- transit delay between two TSAPs;
- transit delay of an Expedited TPDU between two TSAPs;
- transit delay of an Expedited TPDU between two TSAPs for a given T-connection.

Furthermore, a different sequence of specialisation would lead to further characteristics, e.g.:

- transit delay between two TSAPs for a given T-connection.

This approach achieves consistency in two ways. First it gives consistency between different uses of the fundamental concept of the characteristic in different circumstances in that they share a common abstract definition; and second it can be used to give consistency between quite different characteristics where the same specialisations are applied.

Specialisation makes an abstract characteristic more concrete. In any practical application of QOS management a characteristic must be completely specialised, so that it is clear what its values mean. But when developing widely-applicable QOS mechanisms, it may be valuable to work with characteristics at appropriate levels of abstraction. For example, a throughput negotiation mechanism can be defined generically, with the intention that it would be specialised by protocol designers to apply to particular channels and data streams.

This Recommendation | International Standard does not identify or define all the possible specialisations that may be necessary in practice, but does include a subset that may find wide application.

6.2.1.2 Derived characteristics

Some characteristics can be defined as (mathematical) functions of others. These are termed 'derived' characteristics.

One important type of derivation is *statistical*. For example, from the characteristic *throughput* one can derive mean throughput, maximum throughput, minimum throughput, variance of throughput, etc. Technically the statistical derivations are regarded as functions of a random variable that represents the 'base' characteristic from which they are derived.

The following statistical derivations are defined in this Recommendation | International Standard:

- maximum, minimum and range;
- mean;
- variance and standard deviation;
- n-percentile;
- statistical moments.

The precise definitions of these functions are given in Annex B.

The specialisations that apply to a statistical derivation are exactly those that apply to the base characteristic from which it is derived. From *time delay*, for example, one can derive the characteristics *mean transit delay*, *variance of transit delay between two TSAPs*, and so on. Thus, the statistical derivations can be regarded as orthogonal to the specialisations, as illustrated in Figure 6-1.

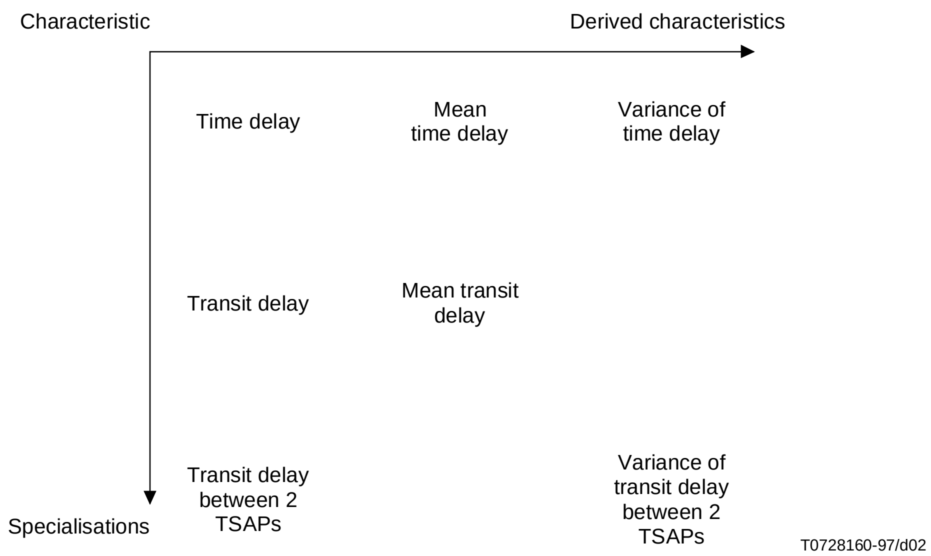


Figure 6-1 – Derived characteristics and specialisations

Other types of derived characteristic can also be defined. In particular, it is possible to define derived characteristics that are mathematical functions of more than one base or derived characteristic. One example is availability, which is a function of maintainability and reliability. It is also possible to derive a characteristic as a function of different specialisations of one or more base characteristics, for example the difference between the throughputs of two channels, or the sum of the transit delays over two sections of a path.

In the case of these more general derivations, the possible specialisations will depend on the base characteristic(s) used and the intended meaning of the derived characteristic.

6.2.2 Descriptive techniques

It will be necessary to define generic QOS characteristics and also some of their specialisations and derived characteristics. The descriptive techniques encompass all of these.

The definition of a QOS characteristic includes the following information:

- a name for the characteristic (NAME);
- a definition that explains its purpose and intended area of application (DEFINED AS);
- a statement of how the characteristic is quantified, giving the units in which values are expressed (QUANTIFIED AS);
- (if a derived characteristic) its STATISTICAL DERIVATION;
- (if a specialised characteristic) its SPECIALISATIONS; and
- (optional) further information.

Ideally, the NAME should be chosen to reflect any specialisation or derivation that has been applied to it. For instance, if some generic characteristic is specialised to apply to transport-connection establishment, its name should express this fact. However, where many levels of specialisation or derivation have been applied, this may not be possible.

It may be useful to define some derivation rules independently of the generic characteristics they may be applied to. Such definitions should include:

- a NAME;
- a NAME COMPONENT for the name of the derived characteristic;
- a definition that explains the meaning of the derivation (DEFINED AS); and
- any further information that may be necessary or desirable.

This is the case for statistical derivations, which may be derived from a wide variety of base characteristics (though not necessarily all) and should be given a common definition.

6.2.3 Specifiers

A number of characteristics may be specialised by applying 'specifiers' that define their application. Three types of specifiers may have wide application: process, location and operation.

6.2.3.1 Process specifier

The process specifier defines the type of events that are concerned. Typically, one or two points P1 and P2 must be considered (what they are is defined under 'location' below) along with the events under consideration (E1, E2, etc.). An example involving a single point is that of 'recovery', where the process is 'failure to provide a service' and the subsequent restoration of service. This leads to the definition of a characteristic 'recovery delay' as the elapsed time between the events E1 (loss of service) and E2 (restoration of service) at a point P.

An example involving two points is that of transit delay, which is defined for some data unit or stream as the elapsed time between two events E1 and E2 where

- E1 is the passage of the first item of data past point P1; and
- E2 is the passage of the last item of data past point P2.

Further process specifiers may be defined to deal with the case of multiple transfers, i.e. to specify that the events E1 and E2 relate to multiple instances of communication. For example, a composite transit delay could be defined for a sequence of data units.

6.2.3.2 Location specifier

The location specifier defines the points P1 and P2 introduced above.

Examples from OSI are:

- (N)-SAP: P1 and P2 are peer (N)-SAPs; and
- SNPA: P1 and P2 are entry and exit SNPAs to a single subnetwork or a chain of subnetworks.

NOTE – Examples of specialised characteristics are (N)-transit delay and SNPA-transit delay.)

6.2.3.3 Operation specifier

The operation specifier defines the types of data that are of concern.

Examples from the communications environment are:

- connection-establishment: establishment and confirmation primitives or PDUs as appropriate;
- connection-release: release and confirmation primitives or PDUs as appropriate;
- CO data transfer: CO SDUs or data PDUs as appropriate;
- CL data transfer: CL SDUs or data PDUs as appropriate.

NOTE – Examples of specialised characteristics in an OSI context are (N)-establishment delay and connectionless-mode (N)-data transit delay.

6.3 QOS characteristics of general importance

6.3.1 Introduction

This subclause describes QOS characteristics of general importance to communications and processing, grouped as follows:

- time-related characteristics;
- coherence characteristics;
- capacity-related characteristics;
- integrity-related characteristics;
- safety-related characteristics;
- security-related characteristics;
- reliability-related characteristics;
- other characteristics.

NOTE – Some QOS characteristics are network dependent, others are related to applications, and some are pervasive across all system elements. Many of the communications-related characteristics identified were initially drawn from CCITT Rec. X.140.

Systems will be implemented with a specific set of characteristics, according to some QOS policy. These sets are considered in 6.4 (QOS categories).

6.3.2 Table of characteristics

Table 6.1 – Table of characteristics

Characteristics group	Name of characteristic	Ref.	Name of specialisation	Ref.
Time-related characteristics	Date/time	6.3.3.1		
	Time delay	6.3.3.2		
	Lifetime	6.3.3.3	Remaining lifetime	6.3.3.4
			Freshness	6.3.3.5
Coherence characteristics	Temporal coherence	6.3.3.6	Temporal data production coherence	6.3.3.7
			Temporal data transmission coherence	6.3.3.8
			Temporal data consumption coherence	6.3.3.9
	Spatial consistency	6.3.3.10	Timeless spatial consistency	6.3.3.11
			Temporal spatial consistency	6.3.3.12
Capacity-related characteristics	Capacity	6.3.3.13	Throughput	6.3.3.14
	Throughput	6.3.3.14	Input user data rate	6.3.3.15
			User information throughput	6.3.3.16
			Application information throughput	6.3.3.17
			Subsystem throughput	6.3.3.18
			Processing capacity	6.3.3.19
	Processing capacity	6.3.3.19	System throughput	6.3.3.20
			Operation loading	6.3.3.21
	Operation loading	6.3.3.21	Association loading	6.3.3.22
			Subsystem loading	6.3.3.23
Integrity-related characteristics	Accuracy	6.3.3.24	Addressing error	6.3.3.25
			Delivery error	6.3.3.26
			Transfer error	6.3.3.27
			Allowable error	6.3.3.28
			Resilience	6.3.3.29
			Transfer integrity	6.3.3.30
			Establishment error	6.3.3.31
			Recovery error	6.3.3.32
			Release error	6.3.3.33
Safety-related characteristics	Safety	6.3.3.34		
Security-related characteristics	Protection	6.3.3.35		
	Access control	6.3.3.36		
	Data protection	6.3.3.37		
	Confidentiality	6.3.3.38		
	Authenticity	6.3.3.39		
Reliability-related characteristics	Availability	6.3.3.40	Channel availability	6.3.3.41
			Connection availability	6.3.3.42
			Processing availability	6.3.3.43
	Reliability	6.3.3.44		
	Fault containment	6.3.3.45		
	Fault tolerance	6.3.3.46		
	Maintainability	6.3.3.47		
Other characteristics	Precedence	6.3.3.48		

6.3.3 Specific characteristics

6.3.3.1 date/time characteristic

DEFINED AS: The absolute time when an event occurs.

QUANTIFIED AS: Any time unit such as minute, second, millisecond, etc. with reference to a known time-origin.

SPECIALISATIONS: Characteristics are specialised from date/time by specifying the particular event, the unit/origin of measurement.

6.3.3.2 time delay characteristic

DEFINED AS: The elapsed time (T2-T1) between two general events E1 and E2, which occur at times T1 and T2.

NOTE 1 – In practical implementations, account may need to be taken of differences between the origins with respect to which T1 and T2 are defined, which may arise from differences in time zones or seasonal times such as summer time and/or synchronisation discrepancies between systems.

QUANTIFIED AS: Any time unit such as minute, second, millisecond, etc.

SPECIALISATIONS: Characteristics are specialised from time delay by specifying the particular events E1 and E2. For example, given locations P1 and P2, transit, request/reply and request/confirm delays can be defined as follows:

transit: E1 is the passage of the first item of data past point P1

E2 is the passage of the last item of data past point P2

request/reply: E1 is the passage of the first item of data past point P1

E2 is the passage of the last item of the related reply past point P1

request/confirm: E1 is the passage of the first item of data past point P1

E2 is the passage of the last item of a related (intermediate) confirmation past point P1

STATISTICAL DERIVATIONS: Any of the statistical derivations defined in 6.2 may be applied to the time delay characteristics defined above.

NOTE 2 – One statistical derivation of particular importance is jitter, which is defined to be the range (minimum to maximum) of a specific time delay over the duration of the communication. For example, jitter is important for data streams in which only small variations in transit delays can be tolerated without significant loss of service to the end user. Typical examples are voice and video data streams, where loss of data is often less damaging than variations in delay.

6.3.3.3 lifetime characteristic

DEFINED AS: The period of time for which the data is valid.

QUANTIFIED AS: Any time unit.

6.3.3.4 remaining lifetime characteristic

DEFINED AS: The time remaining before the data ceases to be valid.

QUANTIFIED AS: Any time unit.

6.3.3.5 freshness (or age of data) characteristic

DEFINED AS: The time since the data was produced.

QUANTIFIED AS: Any time unit.

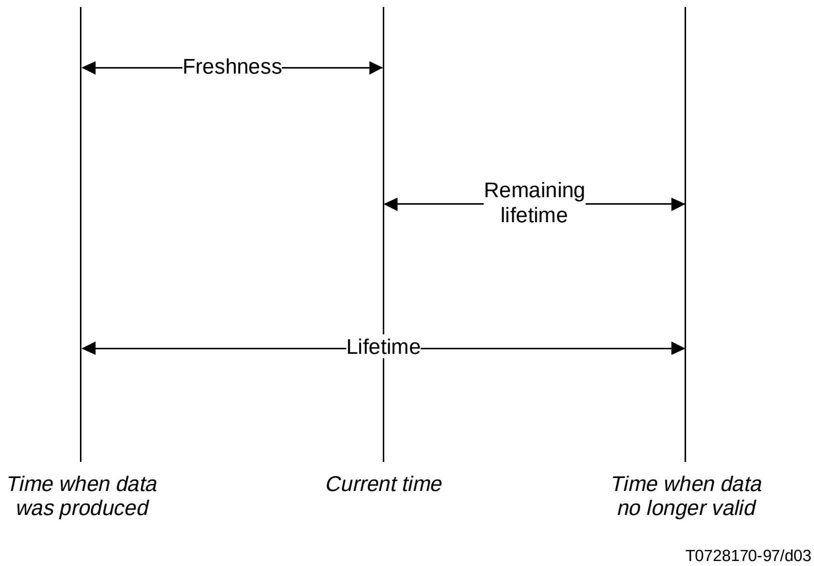


Figure 6-2 – Relationship between freshness and lifetime

6.3.3.6 temporal coherence characteristic

DEFINED AS: Indicates whether an action has been performed on each entity (data item, value, etc.) in a list within a given time window.

QUANTIFIED AS: Boolean value which can take the values 'true' or 'false'.

SPECIALISATIONS: There is a range of possible further specialisations for the temporal coherence characteristic including temporal data production coherence, temporal data transmission coherence and temporal data consumption coherence.

6.3.3.7 temporal data production coherence characteristic

DEFINED AS: Indicates whether the value of each variable in a list has been produced in a given time window.

QUANTIFIED AS: Boolean value which can take the values 'true' or 'false'.

6.3.3.8 temporal data transmission coherence characteristic

DEFINED AS: Indicates whether the value of each variable in a list has been transmitted in a given time window.

QUANTIFIED AS: Boolean value which can take the values 'true' or 'false'.

6.3.3.9 temporal data consumption coherence characteristic

DEFINED AS: Indicates whether the value of each variable in a list has been consumed in a given time window.

QUANTIFIED AS: Boolean value which can take the values 'true' or 'false'.

6.3.3.10 spatial consistency characteristic

DEFINED AS: Indicates whether or not all copies of a duplicated list or multiple copies of a list of variables are identical at a given time or within a given time window.

QUANTIFIED AS: Boolean value which can take the values 'true' or 'false'.

SPECIALISATIONS: There is a range of possible further specialisations for the spatial consistency characteristic, including timeless spatial consistency, temporal spatial consistency, etc.

6.3.3.11 timeless spatial consistency characteristic

DEFINED AS: Spatial consistency where time is not an issue.

QUANTIFIED AS: Boolean value which can take the values 'true' or 'false'.

6.3.3.12 temporal spatial consistency characteristic

DEFINED AS: Spatial consistency within a time window or at a specific time.

QUANTIFIED AS: Boolean value which can take the values 'true' or 'false'.

6.3.3.13 capacity

DEFINED AS: The amount of service that can be provided in a specified period of time.

QUANTIFIED AS: Since the capacity characteristic can be applied to different types of resource, it is quantified using various units. The quantification of capacity also depends on the time unit used for the measurement.

SPECIALISATIONS: Throughput.

6.3.3.14 throughput (communications capacity) characteristic

DEFINED AS: The rate of user data output from a channel averaged over a time interval Δt.

QUANTIFIED AS: A rate such as bits/second or bytes/second

NOTE – Throughput must be a rate; 'user data' should be made explicit in order to differentiate it from control data used to manage the channel (e.g. flow control window parameters); the point at which the rate is defined must be identified; and it is necessary to identify the time interval over which the rate is defined. By varying Δt one can deal with cases where throughput must be maintained essentially constant (Δt small) as well as those where wide variation is permitted so long as a long-term average is met (Δt large), e.g. video streams versus typical packet-switched data.

Some of the 'specifiers' defined for time delay can be applied to throughput too. 'Process' can be used to distinguish between bits/bytes or packets/frames/cells; 'location' can define the channel, e.g. OSI (N)-SAP to peer (N)-SAP; 'operation' can be the type of data in question.

If it is important to ensure that queuing or other indeterminate delays do not occur when user information is transferred, then users need to be able to request their maximum user information throughput for each class of traffic (i.e. degree of urgency) so that they can ensure that determinism is not degraded. Where throughput is affected by load in the communications system, or parts of the system such as the communications stacks, or in the network, it is essential for some users to be able to quantify the effects of loading on the throughput.

SPECIALISATIONS: There is a very wide range of possible further specialisations for the communications throughput characteristic including, input user data rate, user information throughput, application information throughput, subsystem throughput, processing capacity, etc.

6.3.3.15 input user data rate characteristic

DEFINED AS: The rate of user data input to a channel averaged over a time interval Δt.

QUANTIFIED AS: A rate such as bits/second or bytes/second.

NOTE – This characteristic is defined for use in circumstances in which the input and output rates to a channel may differ (other than through statistical variation in delay), for example where data-discarding filters are in operation, or data conversion may take place in the channel.

See also the Note to 6.3.3.14.

SPECIALISATIONS: There is a very wide range of possible further specialisations, as for the communications throughput characteristic defined in 6.3.3.14.

6.3.3.16 user information throughput characteristic

DEFINED AS: An amount of user information transferred in a period of time.

QUANTIFIED AS: See Note under 6.3.3.14 above.

6.3.3.17 application information throughput characteristic

DEFINED AS: An amount of data transferred between applications in a period of time.

QUANTIFIED AS: See Note under 6.3.3.14 above.

6.3.3.18 subsystem throughput characteristic

DEFINED AS: An amount of data transferred over all (N)-associations related to an (N)-subsystem in a period of time.

QUANTIFIED AS: See Note under 6.3.3.14 above.

6.3.3.19 processing capacity characteristic

DEFINED AS: The amount of processing able to be performed in a period of time.

QUANTIFIED AS: Instructions/second.

SPECIALISATIONS: There is a very wide range of possible further specialisations for the processing capacity characteristic including system throughput, loading, etc.

6.3.3.20 system throughput characteristic

DEFINED AS: An amount of processing performed in a period of time.

QUANTIFIED AS: Instructions/second.

6.3.3.21 operation loading characteristic

DEFINED AS: Proportion of capacity being used in a particular time window.

QUANTIFIED AS: Ratio of capacity used to capacity available.

SPECIALISATIONS: There is a very wide range of possible further specialisations for the loading characteristic including: association loading, subsystem loading, etc.

6.3.3.22 association loading characteristic

DEFINED AS: Proportion of (N)-association capacity being used in a particular time window.

QUANTIFIED AS: Ratio of capacity used to capacity available.

6.3.3.23 subsystem loading characteristic

DEFINED AS: Proportion of (N)-subsystem capacity being used in a particular time window.

QUANTIFIED AS: Ratio of capacity used to capacity available.

6.3.3.24 accuracy characteristic

DEFINED AS: The correctness of an event, a set of events, a condition, or data.

QUANTIFIED AS: Probability.

SPECIALISATIONS: Accuracy is a QOS characteristic of concern to the user, for whom this characteristic refers to the integrity of the user information only. (The integrity of headers and similar protocol control information may be the subject of other characteristics.) The accuracy characteristic is specialised in many ways, including addressing error, delivery error, residual error, etc.

6.3.3.25 addressing error characteristic

DEFINED AS: An incorrect choice of address(es) used for delivery of data.

QUANTIFIED AS: Probability.

6.3.3.26 delivery error characteristic

DEFINED AS: Delivery of data to an incorrect address.

QUANTIFIED AS: Probability.

6.3.3.27 transfer error characteristic

DEFINED AS: The incorrect transmission of an amount of data.

QUANTIFIED AS: Probability.

6.3.3.28 allowable error characteristic

DEFINED AS: The amount of errors that may be considered to be acceptable.

QUANTIFIED AS: Probability.

6.3.3.29 resilience characteristic

DEFINED AS: The ability to recover from errors.

QUANTIFIED AS: Probability.

6.3.3.30 transfer integrity characteristic

DEFINED AS: The amount of data transferred in a time interval without error.

QUANTIFIED AS: Probability.

6.3.3.31 establishment error characteristic

DEFINED AS: Inability to establish, within a specified time window, a connection or association that was requested.

QUANTIFIED AS: Probability.

6.3.3.32 recovery error characteristic

DEFINED AS: Inability to recover from an error condition.

QUANTIFIED AS: Probability.

6.3.3.33 release error characteristic

DEFINED AS: Inability to release, within a specified time window, a connection or association.

QUANTIFIED AS: Probability.

6.3.3.34 safety characteristic

DEFINED AS: The level of safety of an event, an action or a resource.

QUANTIFIED AS: Value or level derived from a safety policy.

6.3.3.35 protection characteristic

DEFINED AS: The security afforded to a resource or to information.

QUANTIFIED AS: Probability of failure of the protection.

NOTE – Protection QOS is the degree to which a service provider attempts to counter security threats using security services.

The handling of protection QOS service parameters is a local matter controlled according to the security policy in force. Protection QOS is not negotiated between the service users. For an instance of communication, a service user may indicate its protection QOS requirements to the service provider. A service provider may indicate the protection QOS provided on an instance of communication to the service user. The protection QOS provided by the service provider need not be the same as that requested by the service user.

In OSI, any lower layer protocol exchanges between open systems (referred to as 'in band' protocol exchanges) to convey information about the security services to be selected, are carried in a security association protocol which is independent of an instance of communication. This may be carried implicitly by a security label or explicitly by other means.

For further information on the provision of security in the lower layers and the handling of protection QOS, see ITU-T Rec. X.802 | ISO/IEC TR 13594 – Lower Layer Security Guidelines; and for the upper layers, ITU-T Rec. X.803 | ISO/IEC 10745 – Upper Layers Security Model.

6.3.3.36 access control characteristic

DEFINED AS: Protection against unauthorized access to a resource.

QUANTIFIED AS: Value or level derived from an access control policy.

6.3.3.37 data protection characteristic

DEFINED AS: Protection against unauthorized access to data.

QUANTIFIED AS: Value or level derived from an integrity policy.

6.3.3.38 confidentiality characteristic

DEFINED AS: Protection against unauthorized viewing of data.

QUANTIFIED AS: Value or level derived from a confidentiality policy.

6.3.3.39 authenticity characteristic

DEFINED AS: Protection for mutual authentication and data origin authentication.

QUANTIFIED AS: Value or level derived from an authentication policy.

6.3.3.40 availability characteristic

DEFINED AS: The proportion of agreed service time that satisfactory service is available.

"Agreed service time" means the aggregate time over which it is agreed between service-users and service-provider that service is to be provided.

QUANTIFIED AS: A number in the range (0-1).

For a simple system which has no fault tolerance (or no redundancy) availability can be a simple function of reliability and maintainability:

$$A = MTBF/(MTBF + MTTR)$$

Scheduled maintenance and/or replacement policies can lead to improved levels of availability.

For some applications, the availability requirement is specified over a finite time interval (e.g. 0.99 over a 30 day period). This formulation permits a design where repair or maintenance is not possible during the operational period; the availability is then a function solely of the system reliability.

In more complex systems, the availability requirement can be met even though some elements of the system have failed, by the provision of some degree of redundancy (e.g. a communications service may use an alternate routing in a communications network).

In some cases it will be sufficient to specify availability, and not reliability or maintainability. For some systems it may however be necessary to define a limit to the length of time for which a system can be out of service, in which case the 'downtime' or maintainability characteristic must be specified as well as the availability. Since availability, reliability and maintainability are related, it should never be necessary to specify all three.

NOTE – This is a simplification of the definition in CCITT Rec. X.140, and demands a contractual understanding of what constitutes 'satisfactory'.

SPECIALISATIONS: A number of specialisations of the availability characteristic are possible, including channel availability, connection availability, etc.

6.3.3.41 channel availability characteristic

DEFINED AS: The proportion of agreed service time that the communications channel is available.

QUANTIFIED AS: A number in the range (0-1).

6.3.3.42 connection availability characteristic

DEFINED AS: The proportion of agreed service time that a connection is available.

QUANTIFIED AS: A number in the range (0-1).

6.3.3.43 processing availability characteristic

DEFINED AS: The proportion of agreed service time that a process is available.

QUANTIFIED AS: A number in the range (0-1).

6.3.3.44 reliability characteristic

DEFINED AS: The Mean Time Between Failures (MTBF) to maintain a defined QOS requirement.

QUANTIFIED AS: Any time unit.

6.3.3.45 fault containment characteristic

DEFINED AS: Ability to operate in the presence of one or more errors/faults.

QUANTIFIED AS:

6.3.3.46 fault tolerance characteristic

DEFINED AS: Ability to minimize the impact of component error/failure.

Under assumptions regarding the independence of individual failures and the statistical behaviour of the failures, these characteristics can be related. The reliability of a system is determined by the engineering quality of the hardware and software design, the quality of the components used and the environment in which the system is operated. In principle, the reliability of a system is fixed at design time, but may subsequently be affected by the operational conditions.

QUANTIFIED AS:

6.3.3.47 maintainability characteristic

DEFINED AS: The duration of any continuous period for which satisfactory, or tolerable, service is not available, related to some observation period.

The maintainability of a system is essentially fixed during the design process although small changes may be possible subsequently through maintainer training, alternative upkeep and support strategies.

QUANTIFIED AS: Mean Time To Repair (MTTR) – any time unit.

6.3.3.48 precedence characteristic

DEFINED AS: The relative importance of an object or the urgency assigned to an event.

QUANTIFIED AS: Precedence can be quantified in various ways:

- as a rank ordering of a set;
- as a measure relative to some reference; or
- in comparison to some other object or event.

SPECIALISATIONS: A number of specialisations are possible, including event precedence, resource precedence, function precedence, etc.

6.4 Fundamental QOS categories

6.4.1 Introduction

The system and network architectures which have been standardized to date have not fully considered requirements for specific qualities of service; in effect, they are intended primarily for general services. It is now clear that there are a variety of application areas where users have a specific policy which must be applied during communications and processing activities. These policies will lead to the choice of a particular set of QOS characteristics to implement the requirements.

The concept of a *QOS category* has been developed to allow discussion of the requirements of specific user communities as to what qualities of service best categorise their requirements. The qualities of service which are identified for each category are likely to be similar, but each may have a different emphasis. The QOS categories will not be mutually exclusive; rather a specific community may well demand, say, time-critical, secure and highly reliable systems. A number of candidate QOS categories have been suggested. The list is not exhaustive, but includes:

- secure;
- safety critical, i.e. those user environments which require assurances that failures of software and hardware will not affect the ability to correctly provide services;

- time critical;
- highly reliable, i.e. those user environments where high reliability is essential. This would include those that require highly reliable components and well as those requiring fault tolerance;
- easy to use;
- extensible/flexible;
- monitorable/auditable/testable.

It is intended that the characteristics relevant to each category will be identified by standards groups in each category area.

6.4.2 Time-critical

The time-critical category consists of those user environments which place emphasis on time-related characteristics.

NOTE – Depending on the applications, various aspects may be of importance, like events occurring in a given time window, or coherence of multiple copies of data within a given timeframe, etc.

For some traffic, in time-critical communications, a specific time window exists (and is under user specification or control) during which one or more specialised actions must be completed with a particular level of certainty.

In addition, there is a requirement for temporal and spatial coherence in time-critical data to be supported in TCCS.

For time-critical communications, the underlying architecture must support the user requirements in state-driven, event-driven and mixed systems and must allow time-critical and non time-critical traffic to coexist on the same application association.

7 QOS management

7.1 Introduction

A number of functions are used to manage the quality of the services provided in a system in order to meet the needs of users and applications. The term *QOS management function* is used to identify such functions.

QMFs may require many different types of action to be performed: negotiation, admission control and monitoring, for example. It is therefore useful to regard them as composed of a number of smaller elements, termed *QOS mechanisms*, which can be specified independently. QMFs are then described as being performed by one or more entities operating *QOS mechanisms* in sequence or in parallel.

This clause defines a general structure for QMFs in terms of the phases of the QOS activity in which they are performed and the types of QOS mechanism that are brought into play. It defines some general concepts and terminology. A number of generic QOS mechanisms are identified in clauses 8 and 9. Annex A defines a model of QOS as applied to the specific case of OSI.

Specific QMFs that are defined to meet identified application or user needs will differ in the types of function they provide, the QOS characteristics they affect, the QOS mechanisms they make use of, the entities that perform them, the phases during which they are performed, and so on. QOS mechanisms are defined to perform or support a number of types of activity related to QOS, including:

- QOS establishment;
- QOS monitoring;
- QOS alert;
- QOS maintenance;
- QOS control;
- QOS enquiry.

NOTE – This Recommendation | International Standard places no obligation on network providers to monitor QOS at any time.

7.1.1 Phases of QOS activity

QOS-related activities can be categorised into three phases:

- Prediction phase: The purpose of this phase is to predict aspects of system behaviour so that entities can initiate QOS mechanisms appropriately. In this phase entities will typically make QOS enquiries, and the subjects of the enquiries could include, for example, the current loading of systems elements or previous levels of QOS achieved;
- Establishment phase: The purpose of this phase is to create the conditions so that desired values of QOS characteristics are achieved for some systems activity before that activity occurs. In this phase entities may express requirements for QOS, enter into negotiations or re-negotiations, make agreements on the QOS to be delivered and the actions to be performed if it degrades, and initiate mechanisms that will be needed during the operational phase; and
- Operational phase: The purpose of this phase is to honour the agreements made during the establishment phase, or to take appropriate action of that is not possible. In this phase entities perform QOS monitoring, QOS maintenance and/or QOS enquiries.

The division into phases is not intended to suggest that the activities performed as part of a QMF can always be split into a simple time-sequence of prediction, establishment and operation: its purpose is to classify QOS mechanisms. In practical cases, different entities supporting a single activity may require to be in different phases of QOS activity at the same time. For example, an establishment phase that requires communications at one layer, to support negotiation for example, may require that the operational phase has been entered at lower layers.

Also, a QMF does not necessarily require activity in all three phases. In simple 'best-efforts' QOS, for example, only the establishment phase is used, the operational phase being null as far as QOS is concerned. Furthermore, if it is necessary to modify or re-negotiate QOS requirements while an activity is in progress, a subsequent establishment phase can be entered in parallel with the existing operational phase. (This approach avoids the need to define a 'control' phase that would use exactly the same mechanisms as the establishment phase.)

7.1.2 QOS information

QMFs and mechanisms use and generate information relating to QOS, which is termed *QOS information*. QOS information is independently classified according to its use and according to its meaning, as follows:

•	It is classified into	QOS context	when retained in an entity.
		QOS parameters	when conveyed between entities.
•	It is classified into	QOS requirements	if it expresses a requirement for QOS.
		QOS data	if it does not.

7.1.3 QOS interactions

QMFs may be initiated in three ways. They may be:

- always operational;
- requested by the user of a service, such as an application-process that wishes to ensure, for example, that a certain level of throughput can be provided; or
- instigated by a third party, such as a remote systems manager or a local management process.

Similarly, QOS mechanisms may be initiated by service-users or third parties. QOS mechanisms may be initiated by service-providers, for example, when QOS degradation is detected.

In the particular case of communications between two service-users supported by a service-provider, this leads to a set of interaction requirements between the entities participating in the communications, and a number of potential interactions are identified, together with some examples of their use:

- service-user(s) to service-user(s): to agree on the QOS at which they will operate and the QOS required of the underlying service;
- service-user(s) to service-provider: to request the QOS to be delivered by the service, or to modify the existing QOS;
- service-provider to service-user(s): to convey responses (and indications) to service requests, and to convey monitoring information pertinent to the QOS characteristics of the service;

- service-user(s) to third party(s): to establish or enquire about the QOS environment, and to request monitoring information pertinent to the QOS environment or to a specific interaction;

- third party(s) to service-user(s): to inform a service-user of information pertinent to the QOS environment or a specific interaction;

- service-provider to third party(s): to request assistance in meeting the QOS requirements of service-users;

- third party(s) to service-provider: to respond to such requests.

7.1.4 Use of management mechanisms

In any of the phases, QOS activities may be initiated and performed:

- wholly by entities that participate in the normal course of systems activity;

- in part by management mechanisms, which are outside the scope of this Recommendation | International Standard. (For example, in OSI, this would require OSI layer management or OSI systems management, or a combination of both.)

7.2 Prediction phase

The prediction phase is concerned with establishing the QOS context by making relevant enquiries and performing analysis to predict the QOS characteristics of the system.

From this, it is possible to calculate any potential perturbation if specific actions are taken, to establish the appropriate levels for QOS parameters and to check that requests will not conflict with any admission control policies.

7.3 Establishment phase

7.3.1 Introduction

During the establishment phase, it is necessary for the parties concerned to agree upon the QOS requirements that are to be met in the subsequent systems activity, and to initialise mechanisms to support the operational phase.

This QOS Framework places no obligations on the parties to reach agreements of any particular kinds relating to any particular QOS characteristics: it is for the parties concerned to identify their QOS requirements and to seek to reach agreements appropriately.

An establishment phase is initiated by a service-user or third party that wishes to establish (or re-establish or modify) requirements on one or more QOS characteristics of the activity concerned. The initiator expresses its requirements in terms of QOS requirement parameters conveyed to the appropriate other entity or entities. In some cases a requirement is simply imposed on the other entities, but in others this is the start of a negotiation process that leads, if successful, to an agreement between all the entities concerned on what levels of QOS are to be offered for the characteristics in question and on what actions are to be taken to monitor or maintain QOS, or to signal changes in what is being achieved.

The process of agreeing QOS requirements is described in 7.3.2. It includes the definition of the semantics of QOS requirement parameters relating to negotiations and the associated actions. The agreements made during the establishment phase can result in specific activities being required during the operational phase. There is a relationship between the ability or willingness of a service-provider to perform certain activities and the level of QOS agreement into which that service-provider can enter in response to a QOS requirements. In particular, agreements between a service-provider and its service-users which are at the compulsory or guaranteed levels of agreement (see 7.3.2.4) place a requirement on that service-provider to perform operational phase activities of monitoring and/or maintenance (see 7.4). A number of negotiation mechanisms are identified in 8.2.

The initialisation activities that may be performed during the establishment phase are described in 7.3.2.5. They include the allocation/reservation of resources to the users or providers of the service, and the initiation of mechanisms needed to support the operational phase, such as monitoring.

7.3.2 Agreeing QOS requirements

7.3.2.1 QOS requirement parameter semantics

QOS requirement parameters may have complex semantics, including:

- one or more values of one or more QOS characteristics;
- the role that the value plays in QOS establishment, which may involve negotiation: this role may be:
  - an upper or lower limit;
  - an upper or lower threshold;
  - an operating target;
  - an ancillary parameter, such as a bound of some kind used to limit the possible results of a negotiation;
- actions to be taken as a result of reaching a limit or threshold; and
- the nature of the agreement that the negotiating parties enter into.

In environments where agreements are typically reached for a given set of QOS characteristics, the absence in particular cases of QOS requirements for some of those characteristics may be signalled by the absence of QOS parameters for those characteristics, or by special parameter values having the meaning 'unspecified', or by other means.

It is to be noted that the value of a QOS characteristic may express a statistical as opposed to a deterministic requirement, such as a limit on the mean throughput (calculated over some specified time window), or a target for a probability of error. Such requirements are represented in terms of statistically derived characteristics (as defined, for example, in 6.2.1.2).

Limits and thresholds with which actions are associated are termed 'trigger points'. During QOS establishment it may be necessary to impose or negotiate a number of trigger points, together with an operating target, as illustrated in Figure 7-1 and further to determine the actions associated with them and the nature of the agreement that the parties enter into.

The roles, actions and agreements are described in more detail below.

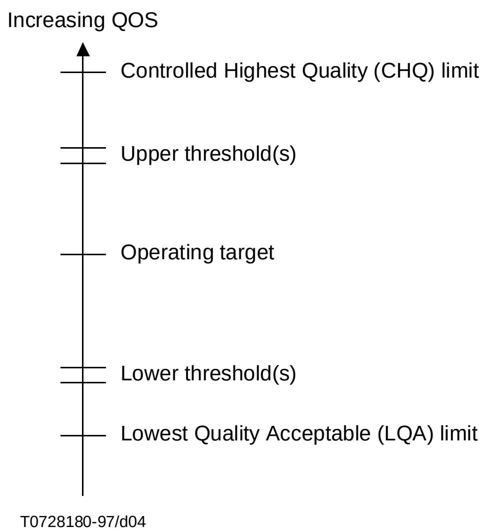


Figure 7-1 – QOS thresholds and limit

7.3.2.2 QOS trigger points and operating targets

Figure 7-1 illustrates the possible relationships between the values associated with limits, thresholds and the operating target. A particular QOS requirement may include one or more of such values.

NOTE – This Recommendation | International Standard places no obligation on any system or communications service to implement any or all of these facilities.

They are described in terms of 'high quality' and 'low quality', rather than numerical values, because for some characteristics high numerical values may correspond to high quality, whereas for others, such as transit delay, high numerical values may correspond to low quality.

A QOS operating target is a negotiated or imposed level at or near which QOS is agreed to be maintained.

There are two types of limit:

- a *Lowest Quality Acceptable* (LQA) limit, below which the QOS 'should not' fall;
- a *Controlled Highest Quality* (CHQ) limit, above which the QOS 'should not' go.

The strength of the prohibition against crossing a limit, and the action taken if it should happen, depends on the agreement entered into, as described in 7.3.2.4.

Thresholds are identified points at which specific actions are defined. They differ from limits in that they carry no semantics that they should not be crossed. Each threshold value must represent lower quality than any CHQ that is specified, and higher quality than any LQA that is specified.

The most general threshold has two specified values, 'high quality' and 'low quality', and an action associated with each. The 'high quality' value must represent QOS greater than or equal to the 'low quality' value. The 'action on high quality' is taken when the high-quality threshold is crossed in the increasing direction, and the 'action on low quality' is taken when the low-quality threshold is crossed in the decreasing direction. Typically, one of the actions is a 'clear' action of some kind, and this is usually the action associated with the threshold value closer to the operating target, if there is one. The general threshold is illustrated in Figure 7-2.

Simpler thresholds are also possible. They may be defined as subsets of the general threshold defined above, with fewer actions and/or just one value as opposed to two. In the case of a single-valued threshold, it is necessary to specify whether the single value is to be understood as 'high quality' or 'low quality', so as to determine the direction of crossing it that will trigger the associated action.

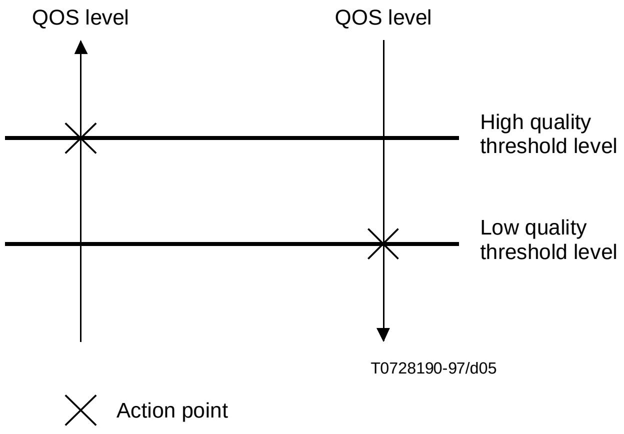


Figure 7-2 – General threshold

7.3.2.3 Actions associated with limits and thresholds

A number of possible actions may be taken by either party upon reaching a limit or threshold:

- take no action (for example, following 'best-efforts' QOS negotiation);
- modify the operation of the service-provider or service-user so as to attempt to keep within limits;
- store a value for future reference (e.g. for enquiry purposes);
- send a signal, for example 'warning' or 'clear', to the service-user or service-provider; or
- abort the service.

NOTE – This Recommendation | International Standard places no obligation on network providers either to monitor achieved QOS or to clear connections or calls when monitored QOS falls below a given limit or threshold.

7.3.2.4 Levels of agreement

The term "level of agreement" is used to describe the actions that the service-provider and/or the service-users agree to take to maintain agreed levels of QOS, as defined in 7.3.2.2 above. The appropriate actions may include:

- no action;
- monitoring the achieved QOS;
- controlling information flow;
- acting to restrict achieved QOS;
- reserving or reallocating resources;
- warning that limits or thresholds are being crossed;
- suspending or aborting the activity;
- suspending or aborting other activities that compete for resources.

Given the range of possible actions, very many possible levels of agreement may be defined. There is no requirement on any particular service-provider to support any of these levels.

This subclause defines three levels of agreement that may be imposed or negotiated, namely:

- Best efforts;
- Compulsory; or
- Guaranteed.

The weakest agreement is that all parties use their *Best Efforts* to meet the users' requirements, but understand that there is no assurance that the agreed QOS will in fact be provided, with no undertaking to monitor the QOS achieved or to take any remedial action should the agreed QOS not be achieved in practice.

In the *Compulsory* level of agreement, the service must be aborted if the achieved QOS degrades below the agreed level; however, the agreed QOS is not guaranteed, and indeed it may be deliberately degraded and the service aborted, for example to satisfy a higher-precedence demand for service.

In the *Guaranteed* level of agreement, the agreed QOS must be guaranteed so that the requested level will be met, barring 'rare' events such as equipment failure. This implies that the service will not be initiated unless it can be maintained within the specified limits.

NOTE – One possible means of implementing a guaranteed level of agreement is the following: once a guaranteed 'Lowest Quality Acceptable' (LQA) value has been accepted, and if other factors prejudice the ability of the system to deliver the guaranteed LQA, then other non-guaranteed instances of provision of service (e.g. compulsory or best efforts) are aborted or additional resources are made available to ensure that the guaranteed LQA is maintained. Only when all non-guaranteed instances of provision of service have been aborted is the guaranteed LQA aborted and the users informed.

Mechanisms that may be used to negotiate agreements of these kinds are subject to separate standardisation activity.

7.3.2.5 Knowledge and perception of QOS agreements

At the end of the establishment phase, it will typically be the case that all participating parties have a common understanding of the precise QOS agreements that have been reached. However, this QOS Framework does not mandate that all negotiation mechanisms must inform all the concerned parties of the common set of agreed values.

In some environments, the values of certain generic QOS characteristics may vary from place to place. For example, in multicast operation, the rate of data reception in one system may be the sum of the rates of data transmitted by other systems; and where filters are employed within a channel, the input data rate may be greater than the output rate. In such cases it is important to define the appropriate specialisations of the QOS characteristics concerned so that the relationships between them can be expressed clearly, and to avoid the kind of description in which QOS characteristics that are really different are represented as 'the same' QOS characteristic, which is perceived by different systems to have different values.

7.3.3 Initialisation mechanisms

The initialisation activities that may be performed during the establishment phase include:

- allocation/reservation of resources to the users or providers of the service;
- initiation of QOS monitoring, for example by setting thresholds;
- setting parameters to be used by the system entities during the operational phase, e.g. window size or frame length.

7.4 Operational phase

The task of the operational phase is to honour the agreements on QOS reached during the establishment phase, so far as is possible, and to initiate appropriate actions where it is not.

The principal activities that may be performed during the operational phase include:

- QOS monitoring, which may be initiated and performed:
  - wholly by entities that participate in the systems activity whose QOS is monitored, in which case it is termed 'local monitoring';
  - in part by management mechanisms, which are outside the scope of this Recommendation | International Standard (in the OSI example, the use of OSI layer management or OSI systems management, or both, would be 'monitoring by OSI Management');
  - filtering.

  NOTE – By their nature, many QOS characteristics cannot be physically, continuously or directly monitored. For example, some characteristics are required to be monitored over extremely long periods in order to observe sufficient events on which to obtain statistical consistency.

- QOS maintenance, which is aimed at maintaining QOS at acceptable levels; QOS maintenance mechanisms include:
  - resource allocation;
  - admission control, i.e. direct control of the rate or times of admission of data at an interface, for example to a communications channel;
  - tuning, i.e. compensatory adjustments of the operation of locally communicating entities, such as those in adjacent OSI layers;
  - synchronization;
  - filtering.
- QOS enquiry, by which entities may request QOS information from other entities.
- QOS alert, by which entities may inform other entities of events that have occurred.

7.5 Supporting services

QMFs may make use of supporting services provided by management entities.

One method of providing general-purpose management is by OSI Systems Management, which can support QOS as follows:

- QOS enquiry and QOS alert between entities in different systems are supported by management operations and notifications. For these purposes it is necessary to define the management capability of the entities in the form of managed object definitions, as defined in the Management Information Model, CCITT Rec. X.720 | ISO/IEC 10165-1 and the Guidelines for the Definition of Managed Objects (GDMO), CCITT Rec. X.722 | ISO/IEC 10165-4.
- Comprehensive monitoring functions are provided by managed objects of the classes defined for Performance Management in the Systems Management Function standards, for example:
  - Metric Objects and Attributes, ITU-T Rec. X.739 | ISO/IEC 10164-11;
  - Summarization Function, ITU-T Rec. X.738 | ISO/IEC 10164-13.
- Scheduling of activities is supported by the Scheduling Function, ITU-T Rec. X.746 | ISO/IEC 10164-15.
- Time synchronisation between systems is provided by the Time Management Function, ITU-T Rec. X.743 | ISO/IEC 10164-20.

QMFs may also make use of supporting services provided for distributed processing, such as those specified in ODP.

8 General QOS mechanisms

8.1 Introduction

This clause classifies a number of QOS mechanisms that can be used to meet the requirements identified in the previous clauses. They are general mechanisms in the sense that they could, in principle, be applied to any QOS characteristic, rather than being specific to any particular QOS characteristics. The management of particular QOS characteristics is the subject of clause 9.

The mechanisms will in general be triggered in one of two ways:

- the input of QOS requirements from another entity, in the form of QOS parameters; this may signal new user requirements, or a change in policy, or a requirement for new operating levels due to changes in the resources available for the provision of a service; or
- the detection of a condition requiring action, for example the fact that a QOS measure crosses some threshold or reaches some limit for which an action is defined (see 8.3.2).

8.2 Prediction phase mechanisms

Prediction phase mechanisms include those for:

- enquiries of historical information on QOS measures which reflect previous levels of QOS achieved;
- analysis of historical information on QOS measures which reflect previous levels of QOS achieved;
- prediction of QOS characteristics in the system (i.e. completion time);
- calculation of potential perturbation if specific QOS requirements are requested and granted;
- evaluation of levels of QOS parameters to be requested in the establishment phase; and
- checking that requests will not conflict with admission control policies.

8.3 Establishment phase mechanisms

8.3.1 Introduction

Establishment phase mechanisms include those for:

- assignment of operating targets, limits or thresholds for particular QOS characteristics, and agreeing the actions to be taken (if any) if the QOS is not maintained; and
- initialising conditions or mechanisms needed during the operational phase.

The assignment of operating targets, limits or thresholds ('levels') for particular QOS characteristics may take place in two ways. The initiator may simply impose a value for a particular level without negotiation by simply informing the corresponding party. Alternatively, a negotiation mechanism may be used. In general, the type of negotiation mechanism that may be used will depend of the semantics of the QOS requirement.

8.3.2 Negotiation mechanisms

Negotiation mechanisms are used to establish operating levels for QOS characteristics and to agree on the actions to be taken if these levels are not maintained. They are typically used in the establishment of communications and play a part in determining which mechanisms are used in the operational phase. They may just require the participation of two interacting parties. For bilateral (peer-to-peer) communications, in the general case they require the participation of three parties (two service-users and a service-provider) and are therefore also described as *three-party* negotiation mechanisms.

In the case of multi-peer communications, QOS characteristics need to be agreed on between the sender, the service-provider and multiple receivers. Depending on the application, a value of a particular characteristic may be agreed on between the sender, the service provider and each individual receiver independently or between the sender, the service provider and all receivers jointly. This leads to a classification of QOS negotiation mechanisms into those for 'connection-wide' and those for 'receiver-selected' negotiation.

In principle, any characteristic may be subject to either style of negotiation, depending on the particular application that is to be supported. Throughput, for example, might be negotiated ‘connection-wide’ where it is important that all parties should receive the data transmitted by a single party without loss. On the other hand, if loss of data can be tolerated, throughput could be negotiated separately for each receiver by ‘receiver-selected’ negotiation. In such circumstances, ‘filters’ that control QOS characteristics by suitable processing (e.g. intelligent frame discarding) can be used during the operational phase to process the sender’s flow to match the needs of individual receivers. This is an example in which negotiation affects the choice of mechanisms used during the operational phase.

Similarly, in many cases transit delay would be negotiated separately for each sender-receiver pair, but for some applications it could be a requirement that synchronisation had to be maintained across all receivers. This would require transit delay to be equally bounded at all receivers and so a connection-wide value would have to be negotiated. In this case, selective buffering could be used to control the QOS achieved.

Negotiation mechanisms are inherently asymmetric in that one party initiates them and one party terminates them. However, this does not imply that the requirements of any party are given greater weight than those of the other(s): it is possible to construct negotiation mechanisms that allow the parties to express their requirements in terms of ranges of acceptable values and derive from those ranges an operating target acceptable to all.

8.3.3 Resource allocation

The establishment phase may invoke mechanisms to allocate resources, such as buffers, circuits, channel capacity and so on, to service-users and/or service-providers. Resources may be allocated in a deterministic fashion, in which case they are reserved for the activity in question, or statistically, in which case they are shared with other activities on the basis that the total available is estimated to be sufficient to meet all the needs, barring rare events. The resources may be allocated once and for all as part of the establishment phase, or they may be subject to re-allocation during the operational phase, for example to counteract a detected degradation of QOS or to support a higher-precedence activity. Mechanisms for resource allocation are subject to separate standardisation.

8.3.4 Initialisation mechanisms

In principle, any of the operational phase mechanisms described below might require initialisation during the establishment phase.

8.4 Operational phase mechanisms

This subclause describes operational phase mechanisms to support:

- QOS monitoring;
- QOS maintenance;
- QOS enquiry;
- QOS alert.

8.4.1 QOS monitoring

QOS monitoring may be initiated and performed:

- Wholly by entities that participate in the normal course of the systems activity whose QOS is monitored, in which case it is termed ‘local monitoring’.
- In part by management mechanisms, which are outside the scope of this Recommendation | International Standard. (In the OSI example, this is termed ‘monitoring by OSI Management’.)

Monitoring mechanisms will need to be invoked by a service-provider if the service-provider has agreed to provide compulsory or guaranteed levels of QOS.

In OSI Management case, monitoring functions are specified in terms of managed object definitions corresponding to the resources that are being monitored. Information can be requested using operations to read attributes, or more generally to perform ‘management actions’ that may be defined specific to the monitoring requirement. Unsolicited information can be generated in the form of notifications.

Monitoring may be scheduled through the use of scheduling packages in the managed object definitions (as defined in Definition of Management Information, CCITT Rec. X.721 | ISO/IEC 10165-2), or the Scheduling Function, ITU-T Rec. X.746 | ISO/IEC 10164-15. Time values may be synchronised between systems through the use of the Time Management Function, ITU-T Rec. X.743 | ISO/IEC 10164-20.

Additional control over monitoring may be achieved by using metric objects (as defined in Metric Objects and Attributes, ITU-T Rec. X.739 | ISO/IEC 10164-11, and other systems management function standards), to report information regularly and to detect and signal threshold-crossing events.

NOTE – This Recommendation | International Standard places no obligation on network providers to monitor QOS at any time.

8.4.2 QOS maintenance

The objective of QOS maintenance is to maintain QOS at acceptable levels. QOS maintenance mechanisms include:

- resource allocation;
- admission control, i.e. direct control of the rate or times of admission of data at an interface, for example to a communications channel;
- output control;
- tuning, i.e. compensatory adjustments of the operation of adjacent layers;
- synchronisation;
- filtering.

8.4.2.1 Resource allocation

As noted in 8.3.3, resource allocation mechanisms are typically invoked during the establishment phase. However, resources may be subject to re-allocation during the operational phase, for example to counteract a detected degradation of QOS or to support a higher-precedence activity. In the former case, they may be coupled with tuning mechanisms (see 8.4.2.4 below).

8.4.2.2 Admission control

Admission control mechanisms limit the acceptance of requests for service from user entities in order to ensure that resources are not overloaded or that existing timing constraints are not disrupted and accepts the users' requests when the requests can, potentially, be met.

8.4.2.3 Output control

Output control mechanisms limit the transfer of data to user entities in order to ensure that their resources are not overloaded or that transfers are strictly timed.

8.4.2.4 Tuning mechanisms

The maintenance of QOS at a certain level is subject to QOS tuning. A control system is needed which takes account of the difference between the desired and the measured QOS and allows feedback to the system to be tuned.

This subclause defines a mechanism for tuning the service provided by QOS to achieve the QOS requested by the user, since QOS may change dynamically due to various reasons (e.g. resource bottlenecks, additional service users, etc.). The implementation of a tuning mechanism is useful either for service designers to realise service guarantees and for external QOS tuning in the case of services having no tuning mechanisms (e.g. to achieve well-defined performance conditions for verification experiments).

The tuning mechanism takes place while the system works and needs real-time calculation.

8.4.2.4.1 System internal tuning

A feedback loop has been introduced for the control of the internal system performance. The service provider will tune a certain internal performance to a particular value to achieve the negotiated QOS.

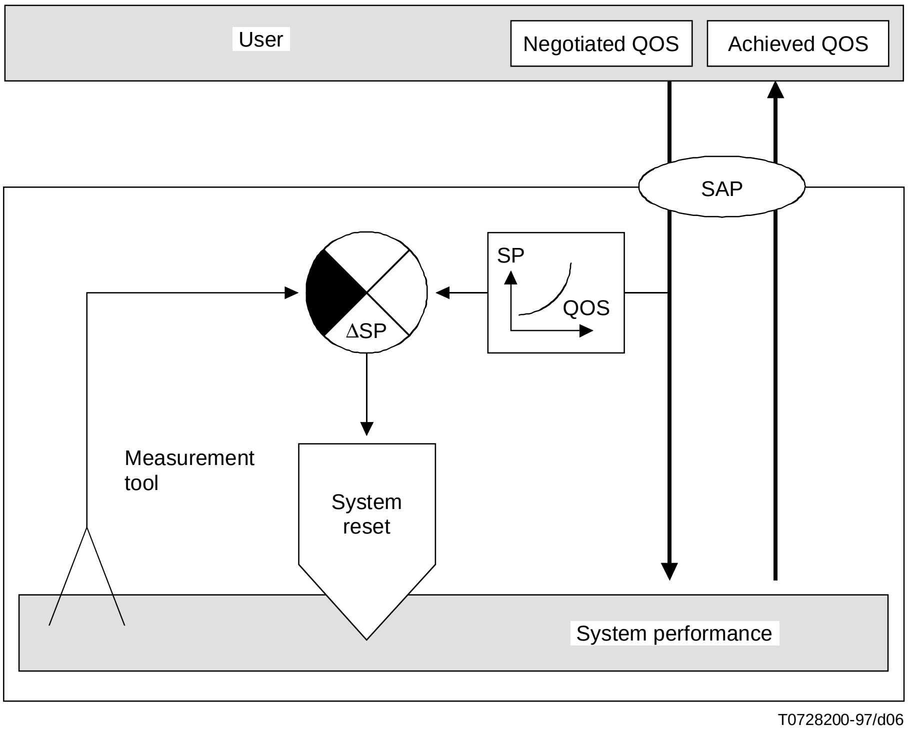
The following operations for system internal tuning have to be performed:

a) calculate required system performance from negotiated QOS;

b) identify (measure) actual system performance;

c) calculate the difference between the required system performance and the measured system performance;

d) reset system facilities to achieve required system performance (negotiated QOS).



SP System performance

Figure 8-1 – System internal tuning mechanism

8.4.2.4.2 System external tuning

System external tuning is applied to systems without system internal tuning. The negotiated QOS will be controlled from the user point of view, i.e. the calculation of the performance difference is done in QOS terms. Therefore, external tuning explicitly serves user requirements. The disadvantage of this mechanism is that QOS to SP and SP to QOS mappings are necessary. This may be provided by relationships established by user experience.

The operations included in the system external tuning mechanism are:

a) measurement of the service-relevant performance parameters (e.g. end system or transfer service);

b) translation to QOS values;

c) calculation of QOS differences; and

d) influence of QOS-related components (end system or transmission service).

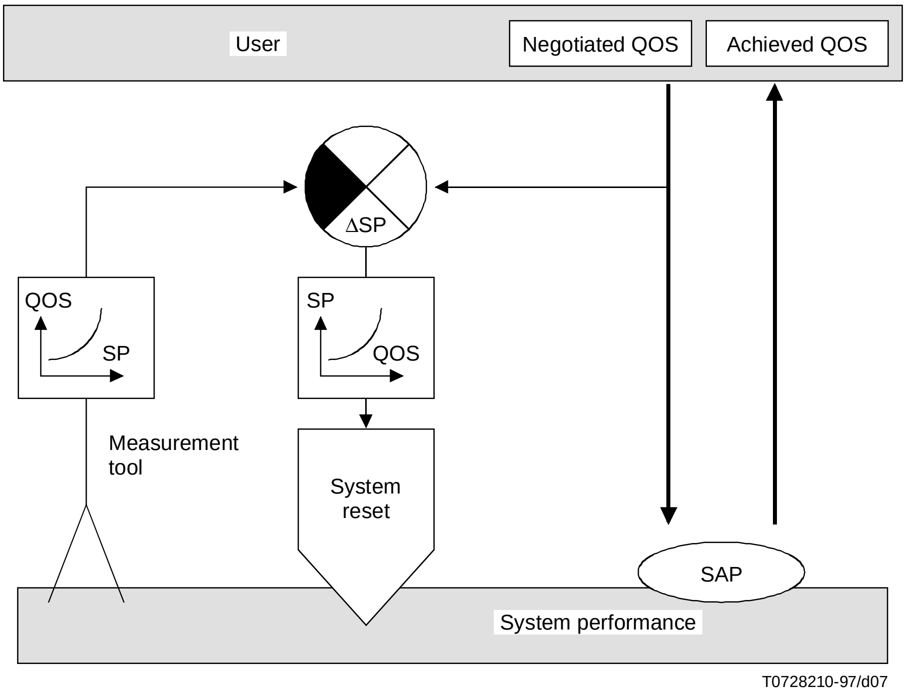
8.4.2.5 Synchronization mechanisms

The operational phase may invoke mechanisms to:

- distribute timing information (e.g. reference clock);
- synchronise actions or events;
- ensure other forms of coherence and consistency are implemented.

8.4.2.6 Filtering mechanisms

Filtering mechanisms are mechanisms that transform data items being transferred in order to alter some of their properties related to QOS and thereby assist in maintaining QOS. Examples include compression and intelligent discarding, both of which can reduce the requirement for raw throughput and therefore enable systems to use lower-capacity channels. Some types of filtering involve trade-offs between a number of QOS properties: discarding, for example, may reduce image quality in order to meet timeliness requirements.



SP System performance

Figure 8-2 – System external tuning mechaism

8.4.3 QOS enquiry and QOS alert mechanisms

QOS enquiry and QOS alert mechanisms are performed by management entities to retrieve management information relating to QOS and to notify changes to QOS characteristics, respectively. Where OSI systems management is used to support QOS enquiry or QOS alerts, it is necessary to define the management capability of the entities in the form of managed object definitions, as defined in the Management Information Model, CCITT Rec. X.720 | ISO/IEC 10165-1 and the Guidelines for the Definition of Managed Objects (GDMO), CCITT Rec. X.722 | ISO/IEC 10165-4.

8.4.4 Dynamic communications scheduling mechanism

Dynamic communications scheduling mechanisms may be required in the operational phase to ensure that the time constraints or timeliness specified by users are respected in the delivery of messages in circumstances in which the behaviour of the network is changing dynamically.

Dynamic communications scheduling mechanisms include those for providing:

- PDU prioritization;
- temporal and spatial coherence; and
- dynamic PDU scheduling.

9 Specific QOS requirements

9.1 QOS requirements related to single QOS characteristics

This subclause discusses a number of QOS requirements specific to particular QOS characteristics.

9.1.1 Time window requirements

Time window requirements are fundamental to the Time Critical QOS category. In general, a time window requirement is a requirement that a particular time delay t shall fall within a given range, that is to say it is a requirement with two parameters t_{min} and t_{max}, such that:

$$t_{min} \leq t \leq t_{max}$$

9.1.2 Throughput requirements

In general a request for a particular throughput is meaningless without information about the offered traffic. Depending on the exact requirement, the additional information may include values of peak, average and burst duration for either or both of bits/bytes and packets/frames/cells. (It should be possible to express these as values of an appropriate throughput characteristic, plus a time period.) Specifically, throughput negotiation mechanisms will often require such additional information to be carried and processed appropriately.

9.1.3 Latent failure detection requirement

Latent failure detection is a requirement fundamental to a wide variety of system categories. It relates to detection mechanisms associated with characteristics which have a bearing on the overall system reliability, stability, etc. By measuring selected characteristics and comparing the results with pre-determined thresholds, latent failure detection may be performed. Having detected latent failures, QOS mechanisms are invoked to perform corrective action.

Latent failure detection may involve a number of mechanisms:

a) Monitoring and analysis of potential perturbation mechanism;

b) Early warning mechanism;

c) Admission control mechanism; and/or

d) Output control mechanism.

9.1.4 Indication of potential perturbation

It is a requirement that any perturbation of a system due to the addition of, or change to, an application process be indicated so that appropriate action may be taken.

Specifically, prediction mechanisms shall be defined and the interactions of such mechanisms with the various elements of the QOS model (see clauses 5 and 6) will need to be appropriately expressed.

The Perturbation Prediction Function predicts potential perturbation and indicates the degree of the potential problems.

9.2 QOS requirements related to multiple characteristics

The most general requirement for QOS is a combination of QOS requirements for individual QOS characteristics, where the combination is tailored to the particular requirements of the application. To assist in specifying such combinations, this subclause identifies the following common cases:

- bulk data: high throughput, low error rate;
- interactive: low delay, low error rate;
- isochronous: high throughput, constant delay;
- time sensitive: constant delay, fixed throughput.

10 QOS verification

10.1 Introduction and stages

QOS verification is the process of comparison of requested QOS with observed and measured QOS in one or more of the design, implementation, testing and operational phases of a given service.

The lifecycle of an implementation of a service is divided into three major stages, in which each stage involves particular tasks needed for QOS verification. In the first stage, the service design stage, it may be a requirement to verify that a given service design specification fulfils the desired QOS requirements as expressed by the service users. In the second, the testing stage, the conformance verification of a service implementation to its specification could also include the measurement and control of System Performance (SP) parameters. During the final stage, service operation, there could be on-line verification that the actual QOS provided to the service users matches the QOS agreed within the service contract. For this task to be accomplished, the service under consideration would have to implement mechanisms for real-time monitoring, managing and tuning of its QOS.

NOTE 1 – There is no obligation that a given service has to provide such mechanisms. The service user is responsible for determining whether that service has such a capability.

NOTE 2 – The service lifecycle stages described above do not correspond to, and should not be confused with, the phases of QOS activity defined in clause 7.

10.2 QOS verification concepts

Achieved QOS is dependent on the performance of the (underlying) system/network used. Therefore any QOS statements from service suppliers depend, in turn, on the specific SP of the service.

QOS verification includes the comparison of requested (agreed) and measured (calculated) QOS. A *QOS verification statement* is formulated as follows:

IF (SP constraints) THEN (QOS requirements)

and means: all QOS measures conveyed as parameter values (or parameter ranges or mean values) are to be verified after the network has been controlled according to the SP constraints specified (see Figure 10-1). SP constraints are Boolean functions which take on the value *true* if QOS measures are found to satisfy user-specified relationships.

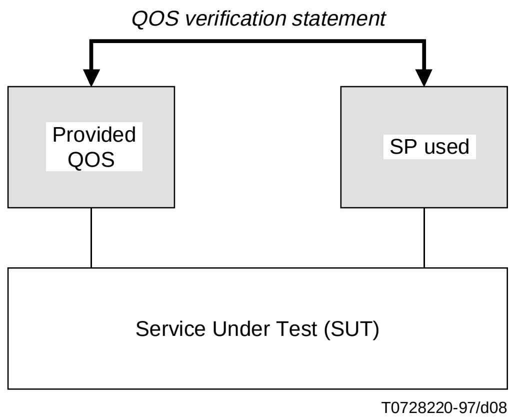


Figure 10-1 – QOS verification statement

11 Conformance, consistency and compliance

11.1 Conformance and the inter-relationship of standards

In the information technology environment, conformance is a property which is claimed by suppliers of implementations of particular standards (i.e. protocol standards or other specifications). This QOS Framework is not itself a specification to which an implementation will be expected to conform, but it is intended to be used as a reference text on QOS by such specifications. Hence, there is no conformance to this Framework.

Relationships between standards themselves are of a different kind and are described using a different terminology. Thus, the nature of the relationship between this QOS Framework and service, protocol and other standards which reference it is such that conventional clauses of conformance do not satisfy the need. The relationship is therefore expressed using the concepts of consistency and compliance[2]. The definitions and statements below utilise the definitions of the approved commentary. Since the relationships between this framework and other frameworks, models, services and protocols which reference it are relationships of consistency and conformity as described below, there is no conformance to this framework. Any verification will be carried out by examination of ITU-T Recommendations | International Standards which reference it.

11.2 Definitions

11.2.1 Consistency

A ‘referencing’ ITU-T Recommendation | International Standard is said to be consistent with a ‘referenced’ ITU-T Recommendation | International Standard if the reference in the ‘referencing’ document does not alter the meaning of the ‘referenced’ standard.

---

[2] These are derived from the approved interpretations of the OSI Basic Reference Model (see ITU-T Rec. X.200 | ISO/IEC 7498-1).

11.2.2 Compliance

A 'referencing' ITU-T Recommendation | International Standard is said to comply with the applicable requirements of a 'referenced' ITU-T Recommendation | International Standard if the following are true:

a) the 'referenced' ITU-T Recommendation | International Standard specifies requirements (using the verb 'shall') which are applicable to the type of ITU-T Recommendation | International Standard of which the 'referencing' ITU-T Recommendation | International Standard is an instance; and

b) the 'referenced' ITU-T Recommendation | International Standard includes a compliance clause to clarify which requirements apply to the type of ITU-T Recommendation | International Standard of which the 'referencing' ITU-T Recommendation | International Standard is an instance,

   i) either the 'referencing' ITU-T Recommendation | International Standard contains a claim of compliance to the 'referenced' ITU-T Recommendation | International Standard; or

   ii) it is possible by inspection of the 'referencing' ITU-T Recommendation | International Standard to verify that the applicable requirements have been fulfilled.

11.3 Application of consistency and compliance requirements

11.3.1 General

The subclauses below set out how ITU-T Recommendations | International Standards which reference this QOS Framework may apply the concepts of consistency and compliance defined above. Statements of consistency and/or compliance with this QOS Framework should be present in referencing ITU-T Recommendations | International Standards in order to minimise inconsistency and ambiguity.

11.3.2 Consistency

An architecture, framework, multi-layer model, single layer model, service definition or protocol specification which claims consistency with this QOS Framework and other QOS ITU-T Recommendations | International Standards which extend this QOS Framework shall state the applicable elements of those set out in the list below:

'This {architecture, multi-layer model, single layer model, service definition or protocol specification}[3]:

a) uses the concepts established by the QOS Framework with identical definitions and terminology for the following terms…;

b) extends the concepts established by the QOS Framework for the following terms…;

c) defines the following concepts…'

The last element in the list above should be used when a term is used in a different sense from that contained in this QOS Framework.

11.3.3 Compliance

11.3.3.1 Compliance of an architecture – Framework or model

An architecture, framework or multi-layer model that is compliant with and which refines this QOS Framework shall state:

'This architecture, framework or multi-layer model is compliant with the QOS Framework in that it describes operations and mechanisms which are defined in the QOS Framework'.

11.3.3.2 Compliance of a protocol specification

A protocol definition that is compliant with this QOS Framework shall state:

'This protocol definition is compliant with the QOS Framework in that it describes functions which pertain to a particular layer as specified in the relevant clause of the QOS Framework'.

11.4 Consistency and compliance with ITU-T Rec. X.200 | ISO/IEC 7498-1

a) This framework is consistent with ITU-T Rec. X.200 | ISO/IEC 7498-1.

b) This framework is compliant with subclauses 5.10, 7.2, 7.4, 7.5 and 7.6 of ITU-T Rec. X.200 | ISO/IEC 7498-1.

[3] The statement would delete those terms which are inapplicable.

Annex A

The model of QOS for OSI

(This annex forms an integral part of this Recommendation | International Standard)

A.1 Introduction

The model of QOS for OSI defines the architectural principles, the concepts and the structures that underlie the provision of QOS in OSI. The model does not itself specify any of the QOS parameters of QOS information that are exchanged when QOS is being established.

Together with the contents of clause 6, the model of QOS for OSI provides the basis for the application to OSI of clauses 7, 8 and 9, in order to specify the provision and management of QOS of OSI communications.

A.2 Architectural principles

The model of QOS for OSI builds on the concepts of the OSI Basic Reference Model, and those of the OSI Management Framework.

The model considers two classes of entities that take part in the management of QOS in open systems:

- system QOS entities (entities which have a system-wide role); and
- layer QOS entities (entities associated with the operation of a particular (N)-subsystem.

The subsystem QOS entities are those that coordinate the response to the requirements imposed on the system. System QOS entities interact with layer QOS entities to monitor and control the performance of the system. In addition, they may implement managed objects, as a means by which systems management entities may interact with the provision of QOS in the system.

The layer QOS entities implement direct control of protocol entities, etc., that are necessary for support of the QOS connections made by the system. In so doing, they respond to the control imposed on them by the system QOS entities, and may negotiate with their layer service's user, and their immediate inferior service provider about the requirements imposed upon them.

Figure A.1 shows the relationship between the system QOS entities and the layer QOS entities. In addition, to their relationship with the (N)-QOS-entities, the system QOS entities also have a relationship with the (N)-service-user and with the (N – 1)-service-provider [(N)-QOS-entities are not (N)-entities in the sense of the OSI Basic Reference Model]. The detail of this relationship is not considered when modelling the (N)-layer.

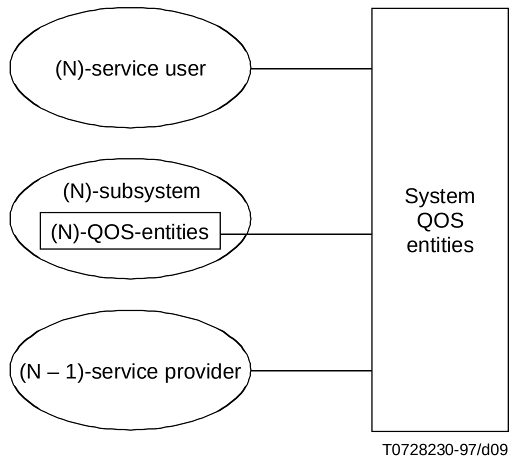


Figure A.1 – The relationship for system to layer QOS entity

Some of the QOS entities may be empty (i.e. they may not perform any function related to QOS management) in open systems conforming to particular sets of OSI standards; there is no requirement that every QOS entity has to be present in every layer. Hence, particular real open systems may be configured to have only a subset of the most general collection of QOS entities described in this subclause.

A.3 Motivation for the provision of QOS

A system that implements measures for specifying, controlling, and monitoring the QOS on its connections is responding to an enterprise requirement for predictability of some aspect or other of its communications. This requirement is represented in system policies; these policies establish the constraints under which all operations of the system may perform.

In turn, these system policies imply layer policies, which apply at each layer of the protocol stack.

Figure A.2 shows the overall model of the way in which provision of a service may respond to the requirements of the service users. There are two overall control inputs to the service provision:

- the user's QOS requirements, which establish initial conditions for the provision of the service; and
- the result of observation of the performance of the communication system, which provides feedback which may cause the network to re-establish the parameters under which it operates.

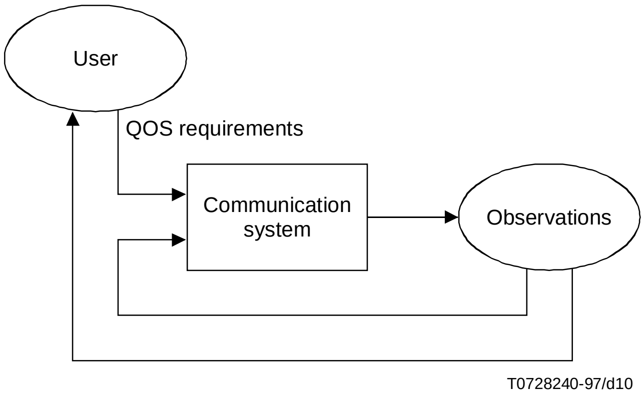


Figure A.2 – The requirement for QOS

A.4 Information flows in the model

A.4.1 Information exchanged

The following kinds of QOS related information elements are conveyed between QOS entities, in QOS parameters:

- QOS requirements: A statement of requirement on one or more QOS characteristics. An example is the set of QOS parameters used in the negotiation of QOS at a particular (N)-SAP.
- QOS data: QOS data is QOS information other than QOS requirements. It includes measures, warnings, requests for information, etc.

The term (N)-QOS-requirements is used to mean QOS requirement parameters that are conveyed across an (N)-service-boundary or between entities within an (N)-subsystem.

A.4.2 Information flow at an (N)-service-boundary

In general, the (N)-service-user and the (N)-service-provider exchange QOS related information. Figures A.3 and A.4 provide a typical example of the flow of QOS requirements at an (N)-service-boundary for unconfirmed and confirmed (N)-service-facilities respectively (the example is taken from existing OSI service standards); however, Figures A.3 and A.4 are not intended to cover all possible cases.

A.4.3 Information flow inside an (N)-subsystem

Figures A.5 and A.6 provide an overview of the flow of QOS requirements between and within (N)-subsystems for both the outgoing and the incoming case (such information flows are described in detail in A.5.7 and A.5.8).

The flow is said to be outgoing when QOS requirements expressed at some point in an (N)-subsystem give rise to either:

- (possibly modified) QOS requirements passed to the (N – 1)-subsystem across the (N – 1)-service-boundary; or
- (possibly modified) QOS requirements being conveyed to the peer (N)-subsystem in (N)-protocol.

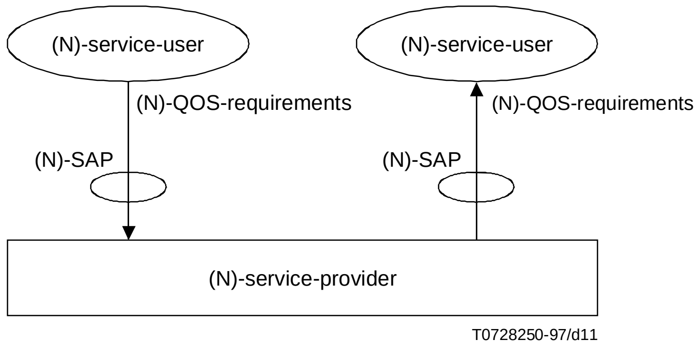


Figure A.3 – Example of flow of QOS requirements in an unconfirmed (N)-service-facility

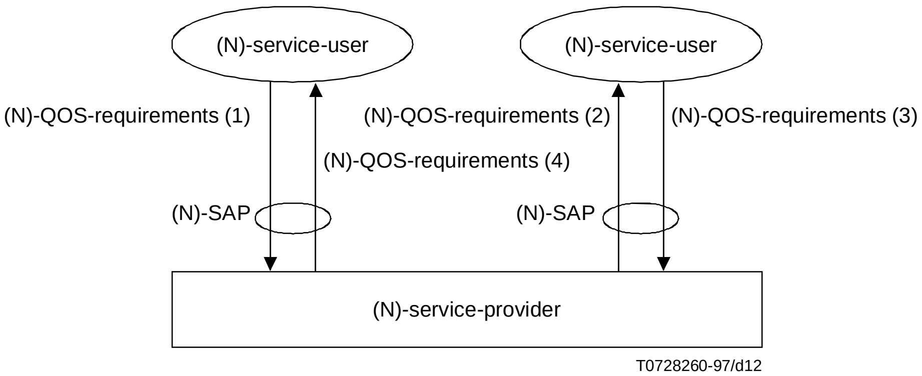


Figure A.4 – Example of flow of QOS requirements in a confirmed (N)-service-facility

Conversely, the flow is said to be incoming when an (N)-subsystem receives QOS requirements either:

- from the (N – 1)-subsystem, across the (N – 1)-service-boundary; or
- contained in (N)-protocol conveyed from a peer (N)-subsystem.

Each of the flows shown in the Figures A.5 and A.6 has a corresponding 'reverse' flow, which occurs when the entity receiving the QOS requirements determines that it cannot meet them. In that case, the responsibility for action passes back to the previous entity in the flow.

Additionally, the (N)-subsystem itself may initiate incoming or outgoing flows without related requirements passing across the (N – 1) or (N)-service-boundary respectively. Such a situation arises when one of the entities in the (N)-subsystem detects a change in achieved QOS of such magnitude that intervention is required. Events of this type can be triggered by QOS monitoring functions in the (N)-subsystem or as a result of interactions with layer or system management. In these cases, Figure A.5 without the upper exchange between the (N)-service-user and the (N)-PCF may apply. Similarly, Figure A.6 can apply without the lower exchange between the (N – 1)-service-provider and the (N)-PCF.

Figures A.5 and A.6 show a decomposition of the (N)-subsystem into (N)-PCF, (N)-QCF and (N)-PE, the roles of which are explained in detail below. This decomposition represents a finer level of description than that used in layer service definitions, and the entire flow between (N)- and (N – 1)-subsystems is regarded as conveyed in (N – 1)-service-primitives, whether it is modelled in this subclause as passing between (N – 1)-PE and (N)-PCF, or between (N – 1)-PCF and (N)-PE.

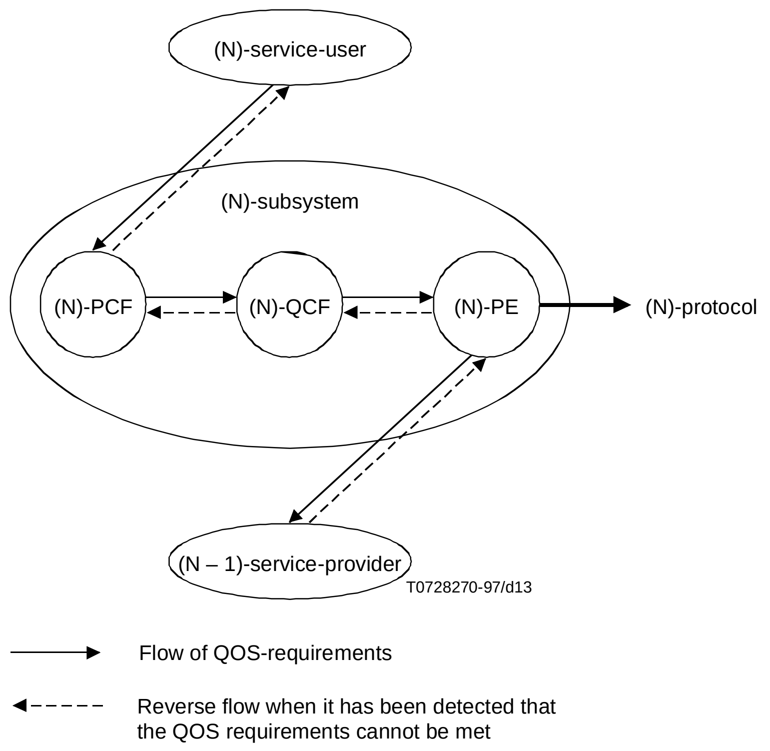


Figure A.5 – Outgoing flow of QOS requirements inside an (N)-subsystem

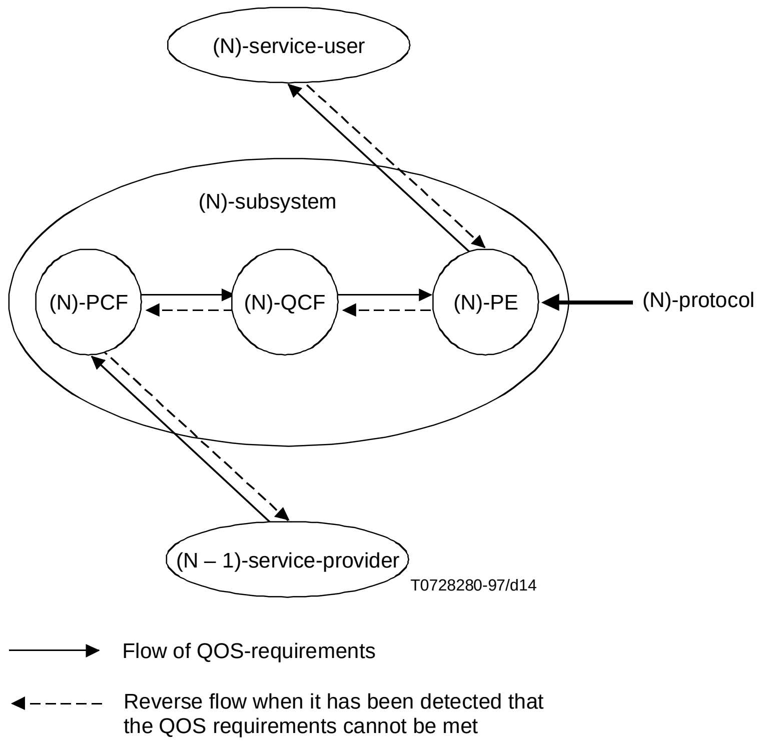


Figure A.6 – Incoming flow of QOS requirements inside an (N)-subsystem

A.5 Layer model of QOS for OSI

A.5.1 Introduction

The layer model of QOS for OSI is only concerned with those aspects of the flows of QOS requirements and QOS data across service boundaries and inside (N)-subsystems that are related to the operation of layer protocols. It models the interactions by which (N)-subsystems at various layers can request, negotiate or declare the QOS at which they and their underlying service-provider are to operate. Further, it models the interactions by which QOS measures, warnings or other information may be exchanged in order to predict, measure or control the QOS achieved. These interactions may occur at any time: they are not restricted to take place during the establishment of a connection or an association, or to coincide with instances of communication, though in many cases they will.

Other information flows (for instance the information flows using systems management) are outside the scope of this Recommendation | International Standard.

In order to perform their functions, in many cases the layer QOS entities may need to reference retained QOS context or access information provided by system QOS entities; however, these are not described in the layer model of QOS for OSI.

NOTE – The current layer model for QOS does not cover cases of multi-peer or multi-cast operation, nor any correlation or synchronisation between different instances of communication. Specifically, it does not cover QOS requirements for transaction processing. A future edition will describe such general OSI capabilities when agreed.

A.5.2 Layering and sub-layering

An open system is regarded as organised into seven (N)-subsystems which can operate and be managed autonomously. Some of the (N)-subsystems are further decomposed following similar principles of independence and autonomy (this is notably the case of layers 2, 3 and 7).

The layer model of QOS for OSI is intended to be applicable to both:

- (N)-subsystems which are not further decomposed, i.e. (N)-subsystems including a single (N)-protocol-entity operating an entire protocol for an OSI layer (e.g. an OSI transport subsystem including the OSI transport protocol entity); and
- autonomous partitions of an (N)-subsystem, i.e. ASOs or the functions inside an open system corresponding to a sub-component of the (N)-layer.

A.5.3 Role of the (N)-service-user

This subclause describes those aspects of the operation of the (N)-service-user that are relevant to QOS management inside the (N)-subsystem.

When exchanging QOS information at a (N)-SAP, the (N)-service-user can assume one of the following roles:

- as a requester, it submits QOS information to the (N)-service-provider and possibly receives related confirmations from the (N)-service-provider;
- as an acceptor, it receives the QOS information delivered by the (N)-service-provider, and possibly submits related responses.

It is to be noted that the (N)-service-user only assumes the role of requester or acceptor for the duration of an (N)-service-facility; hence, in the case of a connection-mode (N)-service, a particular (N)-service-user may assume different roles in different (N)-service-facilities, during the duration of one or multiple (N)-connections.

A.5.4 Role of the (N – 1)-service-provider

This subclause describes those aspects of the operation of the (N – 1)-service-provider that are relevant to QOS management inside the (N)-subsystem.

When exchanging QOS information at an (N – 1)-SAP, the (N – 1)-service-provider performs the following actions:

- it receives QOS information from the (N – 1)-service-user acting as the requester, and possibly delivers related confirmations to that same (N – 1)-service-user;
- it delivers QOS information to the (N – 1)-service-user acting as the acceptor, and possibly receives related responses from that same (N – 1)-service-user.

A.5.5 The roles of the layer QOS entities

A.5.5.1 The (N)-policy-control-function

The (N)-policy-control-function determines policy that is to apply to the operation of the (N)-subsystem. It determines the constraints under which all other decisions of the (N)-subsystems are made. Thus, the (N)-PCF models any actions that must be performed to control the remaining operation of the (N)-subsystem.

For example, the (N)-PCF might apply a security policy. Such policies are beyond the scope of this Recommendation | International Standard, but the (N)-PCF represents the point of interaction at which knowledge of QOS may be used to influence the provision of security, and at which security considerations may affect the provision of QOS. As a consequence of its function, the (N)-PCF may autonomously initiate an incoming or outgoing flow as described in A.5.8 or A.5.7 respectively

It is noted that:

- the application of a policy may cause other activities to be initiated by the (N)-PCF; for example, the application of a security policy may lead to the establishment of a security association;
- in order to perform its functions, the (N)-PCF may need to have access to information provided by system QOS entities, or by systems management entities.

However, these aspects are not part of the layer model of QOS for OSI.

A.5.5.2 The (N)-QOS-control-function

The role of the (N)-QCF is to take account of QOS requirements in selecting the entities that will participate in communications. For example, if a subsystem is capable of operating with any of several (N)-protocol-entities, the (N)-QCF models the influence of QOS considerations on the choice.

At some layers, the (N)-QCF may represent a point of interaction at which knowledge of QOS may be used to influence choice of addresses or routing, and at which addressing or routing considerations may affect the treatment of QOS.

The introduction of the (N)-QCF in the flow of QOS requirements does not imply that there are requirements to select entities in all flows in all layers; in many cases, particularly after the establishment of an instance of communication, its function may be null.

NOTE – The (N)-QCF is currently defined only to participate in the selection of entities (including exerting an influence on addressing and routing). Other possible functions are for further study.

A.5.5.3 The (N)-protocol-entity

The (N)-protocol-entity is responsible for operating the (N)-protocol in order to provide the (N)-service to the (N)-service-users. In particular, it is responsible for negotiating QOS with its peer (N)-protocol-entity (or entities), its (N)-service-user and the (N – 1)-service-provider. As a result of QOS information on which it is acting, the (N)-protocol-entity may also autonomously initiate an incoming or outgoing flow as described in A.5.8 or A.5.7 respectively.

A.5.6 Types of QOS interaction between subsystems

In the descriptions of the flows which follow, the examination, acceptance, rejection or modification of QOS requirements may involve different numbers of parties. Three cases are of general interest:

- involvement of entities in (N)-subsystems in all layers;
- involvement of two (N)-service-users and the entities in the (N)-service-provider;
- involvement of one (N)-service-user and the (N)-service-provider (as in some LAN protocols).

A.5.7 Outgoing flow of QOS requirements

A.5.7.1 Role of the (N)-policy-control-function

The (N)-PCF receives the (N)-QOS requirements submitted by the (N)-service-user and applies specific policies defined for the (N)-subsystem.

As a result of applying policies to the QOS requirements, the (N)-PCF may take one of the following actions:

- reject the QOS requirements and inform the (N)-service-user of the decision;
- accept the QOS requirements, and pass them to the (N)-QCF; or
- modify the QOS requirements, and pass them to the (N)-QCF.

If the (N)-PCF is informed by the (N)-QCF that the QOS requirements are rejected, it may choose to return modified requirements to the (N)-QCF; otherwise it informs the (N)-service-user of the event.

As a result of detected changes in achieved QOS, rejection or modification of the QOS requirements may occur without a specifically related exchange between the (N)-service-user and the (N)-PCF. The (N)-QCF cannot distinguish between this situation and normal events.

A.5.7.2 Role of the (N)-QOS-control-function

The (N)-QCF receives the QOS requirements from the (N)-PCF and performs the following actions:

- it examines the QOS requirements and decides whether the QOS requirements can be met by the operation of an existing (N)-protocol-entity; if such a (N)-protocol-entity exists, the (N)-QCF selects it, otherwise, the (N)-QCF rejects the QOS requirements, and informs the (N)-PCF of the decision;
- after having selected an appropriate (N)-protocol-entity, the (N)-QCF may decide to modify the QOS requirements or to present them to the (N)-protocol-entity in a form which is more suitable to that (N)-protocol-entity;
- it forwards the (possibly modified) QOS requirements to the selected (N)-protocol-entity;
- if it is informed by the (N)-protocol-entity that the QOS requirements have been rejected, it may be able to return modified QOS requirements to the (N)-protocol-entity or to select an alternative (N)-protocol-entity; otherwise it informs the (N)-PCF of the event.

A.5.7.3 Role of the (N)-protocol-entity

The (N)-protocol-entity must check that the QOS requirements imposed on it can be supported. In the course of this, it may need to:

- select protocol features necessary to provide the requested QOS;
- select QOS parameters, and exchange or negotiate them with its peer (N)-protocol-entity;
- pass QOS requirements to the adjacent layer [i.e. to the (N)-service-provider or to the (N)-service-user, as appropriate].

Any of these operations may fail, in which case the requested QOS requirements cannot be supported. In this case, the failure is reported back to the source of the requirements via the (N)-QCF.

Further, in the course of operation of a QOS-controlled association, the (N)-protocol-entity may observe that the achieved QOS is lower than that accepted for the association. In this case, the (N)-protocol-entity may inform the (N)-QCF of the situation, and a further interaction may take place. In so doing, an outgoing flow of QOS requirements may take place without a specifically related exchange between the (N)-service-user and the (N)-PCF.

A.5.8 Incoming flow of QOS requirements

A.5.8.1 Role of the (N)-policy-control-function

The (N)-PCF receives the (N – 1)-QOS-requirements delivered by the (N – 1)-service-provider, and applies specific policies defined for the (N)-subsystem.

As a result of applying policies to the QOS requirements, the (N)-PCF may perform one of the following actions:

- reject the (N – 1)-QOS-requirements and inform the (N – 1)-service-provider;
- accept the (N – 1)-QOS-requirements and pass them to the (N)-QCF; or
- modify the (N – 1)-QOS-requirements and pass them to the (N)-QCF.

If the (N)-PCF is informed by the (N)-QCF that the QOS requirements have been rejected, it may choose to return modified requirements to the (N)-QCF; otherwise it informs the (N – 1)-service-provider of the event.

As a result of detected changes in achieved QOS, the (N)-PCF may modify the (N)-QOS-requirements autonomously and pass them to the (N)-QCF.

A.5.8.2 Role of the (N)-QOS-control-function

The (N)-QCF receives the QOS requirements from the (N)-PCF and performs the following actions:

- it examines the QOS requirements and decides whether the QOS requirements can be met by the operation of an existing (N)-protocol-entity; if such a (N)-protocol-entity exists, the (N)-QCF selects it, otherwise, the (N)-QCF rejects the QOS requirements, and informs the (N)-PCF of the decision;

- after having selected an appropriate (N)-protocol-entity, the (N)-QCF may decide to modify the QOS requirements or to present them to the (N)-protocol-entity in a form which is more suitable to that (N)-protocol-entity;
- it forwards the (possibly modified) QOS requirements to the selected (N)-protocol-entity;
- if it is informed by the (N)-protocol-entity that the QOS requirements have been rejected, it may be able to return modified QOS requirements to the (N)-protocol-entity or to select an alternative (N)-protocol-entity; otherwise it informs the (N)-PCF of the event.

A.5.8.3 Role of the (N)-protocol-entity

The (N)-protocol-entity may receive QOS requirements from either the (N – 1)-service-provider [via the (N)-QCF] or its peer (N)-protocol-entity; these sources may impose requirements separately or simultaneously. In each case, the (N)-protocol-entity may choose to negotiate with the source of the requirements without reference to the (N)-service-user [4].

In any of the cases, after checking that the requirements can be met at all, the (N)-protocol-entity may need to:

- modify the QOS requirements so that they can be used;
- negotiate with the source of the requirements [possibly with the participation of the (N)-service-user].

When this process is complete, the (N)-protocol-entity will deliver QOS requirements to the (N)-service-user; these requirements are derived from those that the (N)-protocol-entity received, after possible modification.

If at any stage the operation fails (because no adequate protocol mechanism is available to support the requirements, because negotiation fails, or because the (N)-service-user rejects the requirements passed to it), the (N)-protocol-entity informs the immediate source of the requirements (i.e. either the (N)-QCF or the peer (N)-protocol-entity) of the failure. In the course of operation of a QOS-controlled association, the (N)-protocol-entity may observe that the achieved QOS is lower than that accepted for the association. In this case, the (N)-protocol-entity may deliver QOS requirements to the (N)-service-user. In so doing, an incoming flow of QOS requirements may take place without a specifically related exchange between the (N – 1)-service-provider and the (N)-PCF.

A.6 System model of QOS in OSI

The foregoing subclauses have described the operation of QOS in OSI as the collaboration of a number of layer QOS entities. This collaboration will, in real open systems, typically be supported by stored information and processing functions that are not specific to individual OSI layers but act to coordinate the whole. Where such information and functions neither engage in external communications themselves nor are otherwise visible to external systems, following OSI practice they are not modelled as entities in open systems but are left to be determined by implementation choice.

However, open systems may contain elements that are not layer-specific but are externally visible and play a role in managing QOS. These include management entities, a system quality control function and a system policy control function.

A.6.1 Management entities

Management entities will commonly be present in open systems to support OSI systems management (and perhaps also OSI layer management). Those supporting OSI systems management may be regarded as falling into two classes: those that provide the management infrastructure, by supporting management communications (CMIP), event-forwarding, logging and so on; and those which use the management infrastructure to perform particular management tasks, by acting as managers or managed resources (modelled by managed objects).

The application of QOS to management communications is a special case of the operation of an OSI layer, and is covered by the treatment in the foregoing subclauses. But systems management can also be used to provide communications paths for QOS information which are asynchronous with the communications whose QOS is being managed and will, in general, involve third parties.

To achieve this, the roles of the elements concerned must be specified in systems management terms, i.e. as managers and managed resources within a defined management environment of managing and managed systems; and the capabilities of the managed resources must be specified in the form of managed object definitions, as defined in CCITT Rec. X.720 | ISO/IEC 10165-1 and CCITT Rec .X.722 | ISO/IEC 10165-4. This applies to any of the elements described in this subclause, including the system quality control function and system policy control functions below.

---

[4] The case of negotiating QOS requirements with a peer (N)-protocol-entity may occur in certain Local Area Network protocols. Since it does not allow the (N)-service-user to participate in the negotiation, it is not recommended for general use.

In addition to the normal use of OSI Systems Management Functions to support general management operations and notifications, the following may be of particular importance for QOS Management:

- Metric Object and Attributes: To provide statistics (means, variances, etc.) of QOS data and to make them available in QOS alerts or responses to QOS enquiries.
- Summarization Function: To assemble a variety of information relating to a particular time window and make it available in QOS alerts or responses to QOS enquiries.
- Time Management Function: To synchronize clocks between systems and support the calculation of time intervals.

A.6.2 System quality control function

The system quality control function provides two capabilities: a system-wide capability for tuning the performance of the various protocol entities that are in operation [in contrast to direct (N)-service-user/(N)-service-provider interactions, as described in 8.4.2.4], and providing coordination of any requirement to modify the behaviour of remote systems via OSI systems management.

A.6.3 System policy control function

The role of the system policy control function is similar to the role of the (N)-PCF in a layer. The inclusion of the system policy control function as an entity recognises that any policy implemented at any particular layer is likely to depend on a policy which has been established for the whole open system and which may require communications with an external policy manager to maintain or vary it.

For example, in the case of a policy concerned with time critical communications, the system policy control function may need access to information concerning not just the single open system, but also other open systems involved in the time critical communication; and in the case of security policies, communications with security managers may need to be supported.

Annex B

Definitions of statistical derivations of characteristics

(This annex does not form a integral part of this Recommendation | International Standard)

This annex defines the statistical derivations listed in 6.2.1.2, namely:

- maximum, minimum and range;
- mean;
- variance and standard deviation;
- n-percentile;
- statistical moments.

The 'base' characteristics to which the statistical derivations are to be applied are regarded as random variables with certain probability distributions, and the statistical derivations are functions of those random variables, as defined below.

The derivations are defined for a real random variable *X*, which has a probability distribution function *F*, where:

$$F(x) = \Pr(X \le x)$$

If *X* has maximum and minimum values, the **range** of *X* is defined by $R(X) = \max(X) - \min(X)$.

The **mean** (or **expectation**) of *X* is most generally defined to be the Lebesgue integral:

$$\mu_x = E(X) = \int x dF(x)$$

Where *X* has a continuous distribution with density function $f(x)$, this becomes:

$$\mu_x = \int x f(x) dx$$

and where *X* has a discrete distribution with probabilities P_i, where $P_i = \Pr(X = x_i)$,

$$\mu_x = \sum x_i p_i$$

The **variance** var(*X*) and **standard deviation** σ_x are defined by:

$$\sigma_x^2 = \mathrm{var}(X) = E(X - \mu_x)^2$$

where

$$E(X - \mu_x)^2 = \int (x - \mu_x)^2 \, dF(x)$$

The **upper n-percentile** U_n is defined for a continuous distribution to be that value which is exceeded with probability *n*% and the **lower n-percentile** L_n to be that value which is exceeded with probability 1 – *n*%.

Various moments can also be defined, though they are probably of limited application. The general **nth central moment** is defined by:

$$C_n = E(X - \mu_x)^n = \int (x - \mu_x)^n \, dF(x)$$

Annex C

Relationships between QOS Recommendations | International Standards and other Recommendations | International Standards

(This annex does not form a integral part of this Recommendation | International Standard)

Figure C.1 shows how the major QOS documents currently relate to other documents.

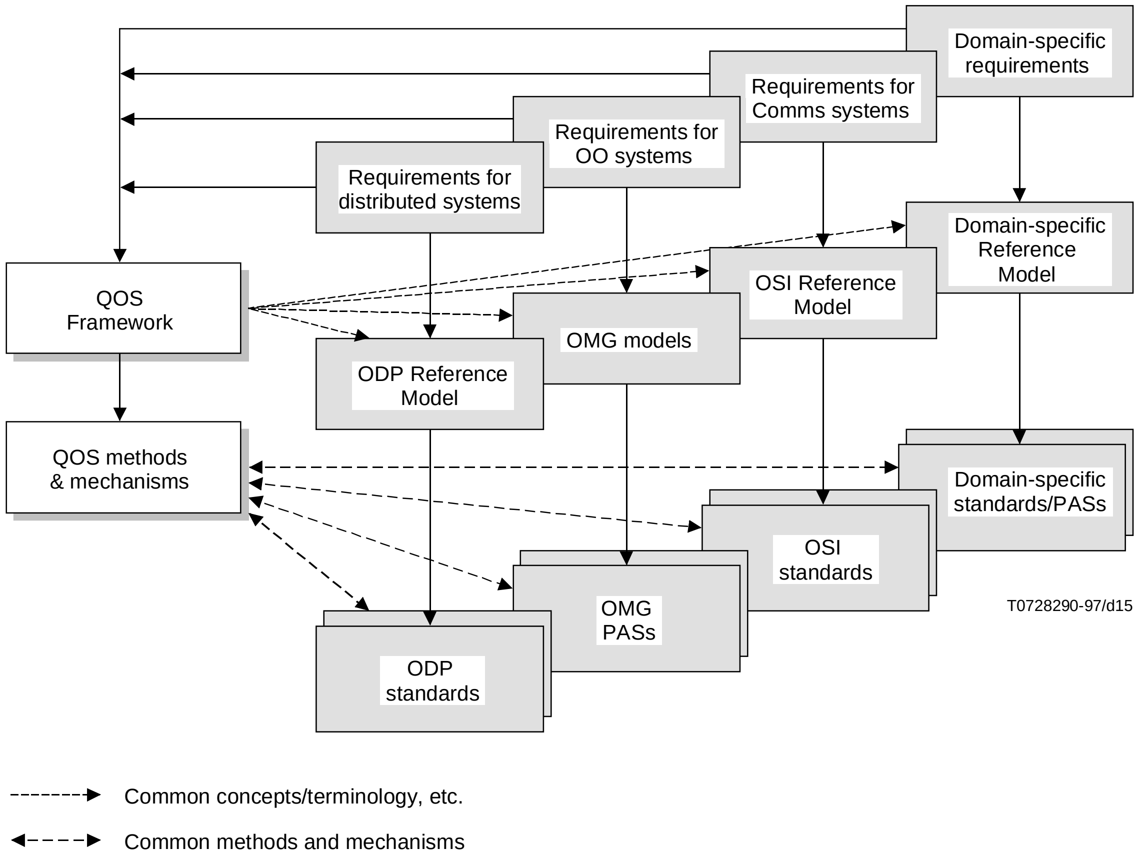


Figure C.1 – Relationship between documents

This **QOS Framework** provides a common basis for the coordinated development and enhancement of a wide range of Recommendations | International Standards that specify or reference Quality of Service (QOS) requirements or mechanisms. It offers a means of developing or enhancing Recommendations | International Standards relating to QOS and provides concepts and terminology that will assist in maintaining the consistency of related Recommendations | International Standards.

The initial work in developing this QOS Framework was done in the context of OSI with the objective of supplementing and clarifying the description of QOS contained in the Basic Reference Model of Open Systems Interconnection (OSI) (ITU-T Rec. X.200 | ISO/IEC 7498-1). It is recognised, however, that management of QOS is important not only in OSI communications but also in a much wider context, and that there is value in encouraging a common approach to QOS that can extend to other communications architectures, to distributed processing in general, and to Open Distributed Processing (ODP) in particular.

Hence this Recommendation | International Standard is structured and written in such as way as to encourage many communities to adopt its approach, concepts, terminology and definitions. Its concepts and terms are defined without reference to any particular architecture, so that they can be adopted and applied by other communities to a variety of architectures and protocols. This general treatment is supplemented by examples from OSI, ODP and elsewhere; thus, it provides a conceptual and functional framework for QOS which allows independent teams of experts to work productively on the development of Recommendations | International Standards.

The **QOS Methods and Mechanisms** (see ITU-T Rec. X.642 | ISO/IEC TR 13243: *Information technology – Quality of Service – Guide to methods and mechanisms*) provides information and guidance on current and proposed approaches to QOS in a wide variety of areas of specification. It includes references to Recommendations | International Standards, other specifications and on-going work related to QOS, as well as definitions of a number of widely-used QOS mechanisms. The references and definitions should be useful to those specifying QOS-enhanced systems and will also serve as the basis for identifying further commonalities.

Annex D

Cost Information

(This annex does not form a integral part of this Recommendation | International Standard)

This annex discusses the issue of 'cost' in relation to Quality of Service.

In many cases, users of services will wish to impose some constraints on the costs that may be incurred in meeting the QOS requirements they express. Thus, information relating to cost may need to be transferred at the same time as QOS information and may need to be taken into account in the process of QOS negotiation. However, the cost of provision of a service may depend on many factors other than the immediate service requirement, for example, contractual arrangements and long-term service usage. For this reason, cost is treated differently from QOS characteristics.

NOTE 1 – This information on cost places no obligation on common carriers or their services to provide any information or take any action.

Although cost may be quantified in terms of currency/unit, it should not necessarily be seen as representing the actual, instantaneous, cost of an event or provision of a service. In many circumstances, a relative indication may be sufficient to enable an appropriate selection to be made between different potential choices. In some interpretations, cost could represent a wider 'business cost' view, including the investment necessary to provide facilities, e.g. memory, storage, processing power, skills, opportunity cost, etc. In other interpretations, when resources needed for an event can be clearly evaluated, cost information may represent the absolute cost of the event or service.

In some instances the cost could be measured as currency/unit but might be more usefully expressed in terms of other resources consumed or denied as a result of the user's QOS requirement.

It must be recognised that communications costs are quite often very difficult to determine accurately prior to the communications, and that the acts of performing cost calculations, monitoring and reporting communications costs may increase the costs significantly. Thus, the ability to provide communications services without performing cost calculations or cost monitoring (especially where simple low cost communications are needed) should be an available option. Also, where cost estimation calculations are performed, costs that are within a narrow range should be considered to have the same value (to avoid 'one-penny price wars' among competitive systems).

The cost of a service is often a function of the QOS options selected and may, in some cases, be calculable by the service user from information supplied by its service provider(s). A user may consume resources which have an implied cost because their use denies some resource (e.g. bandwidth) to other users.

NOTE 2 – The expression of cost of service will vary in precision. The minimum level of expression necessary, where cost information is provided to the user, is that to enable a selection to be made between differing courses of potential action to satisfy requests made for QOS characteristics. This might be accomplished by a parameterisation as simple as low/medium/high. More specific expressions of cost may be made when characteristics which involve many discrete points or a continuum of possibilities are involved. Even when this is the case, it must be recognised that an expression of cost in a QOS context will not be sufficient to express or replace cost factors which involve unrelated or distantly related events in the same enterprise, e.g. bulk corporate discounts, usage thresholds, etc., which are subject to contractual arrangements. It is not possible to capture the impact of such factors in the definitions and representations in this framework.

Annex E

QOS Bibliography

(This annex does not form a integral part of this Recommendation | International Standard)

- ITU-T Recommendation X.642.....[5)] | ISO/IEC TR 13243:.....[5)], *Information technology – Quality of Service – Guide to methods and mechanisms.*
- ITU-T Recommendation X.701 (1997) | ISO/IEC 10040:1998, *Information technology – Open Systems Interconnection – Systems management overview*.
- CCITT Recommendation X.720 (1992) | ISO/IEC 10165-1:1993, *Information technology – Open Systems Interconnection – Structure of management information: Management information model plus Technical Corrigendum 1 (1994).*
- CCITT Recommendation X.721 (1992) | ISO/IEC 10165-2:1992, *Information technology – Open Systems Interconnection – Structure of management information: Definition of management information plus Technical Corrigendum 1 (1994).*
- CCITT Recommendation X.722 (1992) | ISO/IEC 10165-4:1992, *Information technology – Open Systems Interconnection – Structure of management information: Guidelines for the definition of managed objects.*
- CCITT Recommendation X.736 (1992) | ISO/IEC 10164-7:1992, *Information technology – Open Systems Interconnection – Systems management: Security alarm reporting function.*
- ITU-T Recommendation X.738 (1993) | ISO/IEC 10164-13:1995, *Information technology – Open Systems Interconnection – Systems management: Summarization function.*
- ITU-T Recommendation X.739 (1993) | ISO/IEC 10164-11:1994, *Information technology – Open Systems Interconnection – Systems management: Metric objects and attributes.*
- ITU-T Recommendation X.743[6)] | ISO/IEC 10164-20:.....[6)], *Information technology – Open Systems Interconnection – Systems management: Time management function.*
- ITU-T Recommendation X.802 (1995) | ISO/IEC TR 13594:1995, *Information technology – Lower layers security model*.
- ITU-T Recommendation X.803 (1994) | ISO/IEC 10745:1995, *Information technology – Open Systems Interconnection – Upper layers security model*.

[5)] Presently at the stage of draft.

[6)] To be published.

ITU-T RECOMMENDATIONS SERIES

Series A	Organization of the work of the ITU-T
Series B	Means of expression: definitions, symbols, classification
Series C	General telecommunication statistics
Series D	General tariff principles
Series E	Overall network operation, telephone service, service operation and human factors
Series F	Non-telephone telecommunication services
Series G	Transmission systems and media, digital systems and networks
Series H	Audiovisual and multimedia systems
Series I	Integrated services digital network
Series J	Transmission of television, sound programme and other multimedia signals
Series K	Protection against interference
Series L	Construction, installation and protection of cables and other elements of outside plant
Series M	TMN and network maintenance: international transmission systems, telephone circuits, telegraphy, facsimile and leased circuits
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	Telephone transmission quality, telephone installations, local line networks
Series Q	Switching and signalling
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminals for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	**Data networks and open system communication**
Series Y	Global information infrastructure
Series Z	Programming languages