INTERNATIONAL TELECOMMUNICATION UNION

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

X.642

(09/98)

SERIES X: DATA NETWORKS AND OPEN SYSTEM COMMUNICATIONS

OSI networking and system aspects – Quality of service

Information technology – Quality of Service – Guide to methods and mechanisms

ITU-T Recommendation X.642

(Previously CCITT Recommendation)

ITU-T X-SERIES RECOMMENDATIONS

DATA NETWORKS AND OPEN SYSTEM COMMUNICATIONS

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

TECHNICAL REPORT 13243

ITU-T RECOMMENDATION X.642

INFORMATION TECHNOLOGY – QUALITY OF SERVICE – GUIDE TO METHODS AND MECHANISMS

Summary

This Recommendation | Technical Report promotes the use of common methods and mechanisms for managing Quality of Service in information technology and data communications services and protocols. These methods and mechanisms are consistent with the overall Quality of Service Framework given in ITU-T Rec. X.641 | ISO/IEC 13236.

Source

The ITU-T Recommendation X.642 was approved on the 25th of September 1998. The identical text is also published as ISO/IEC Technical Report 13243.

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

This Recommendation | Technical Report is intended to promote the use of common methods and mechanisms for managing Quality of Service (QoS) in a number of different communications and systems environments.

The collaborative ITU-T and ISO/IEC JTC 1 project for coordination and harmonization of QoS-related initiatives is aimed at encouraging the use of the QoS Framework (see ITU-T Rec. X.641 | ISO/IEC 13236), facilitating the use of common QoS methods and mechanisms, and promoting consistency between different applications and systems in their treatment of QoS. The collaboration has been extended to include QoS in Open Distributed Processing. Efforts are being made to promote maximum consistency between this activity and the work on development of specifications for QoS in CORBA-based systems in the Object Management Group (OMG).

Organizations developing methods or mechanisms for QoS management are encouraged to make use of the concepts and terms defined in the QoS Framework. Any developments that may be re-usable in other contexts should also be proposed for reference in this Recommendation | Technical Report by submitting a reference, together with explanatory text, to the ITU-T or JTC 1 Secretariat.

It is expected that the convergence of QoS methods and mechanisms will be achieved in a stepwise fashion, using such submitted material as a basis. In its first edition, this Recommendation | Technical Report identifies and catalogues current standards and other widely available specifications that incorporate definitions of QoS characteristics and QoS methods and mechanisms; and it includes definitions of some methods and mechanisms that are considered to be widely applicable. These methods and mechanisms are derived from those used or under development in information technology standards, and have been formulated in a manner consistent with the QoS Framework, with the objective that they can be applied widely and, if appropriate, standardized. Subsequent editions of this Recommendation | Technical Report are expected to add further methods and mechanisms, likewise formulated in a manner consistent with the QoS Framework. Through this process, harmonization of QoS approaches and usage across a wide range of environments will be achieved.

Since this Recommendation | Technical Report includes methods and mechanisms developed elsewhere, in cases of conflict between definitions in this Recommendation | Technical Report and definitions in the source specifications, the latter have precedence.

TECHNICAL REPORT

ITU-T RECOMMENDATION

INFORMATION TECHNOLOGY – QUALITY OF SERVICE – GUIDE TO METHODS AND MECHANISMS

1 Scope

This Recommendation | Technical Report uses the concepts and terminology of the Quality of Service Framework, ITU-T Rec. X.641 | ISO/IEC 13236. It is intended to support those designing, testing and specifying Information Technology (IT) systems, data communications services and protocols, those defining QoS management functions and QoS mechanisms for particular data environments and technologies, and those engaged in other QoS-related activities such as system testing, by providing a source of reference material on QoS. To do this, it brings together references to methods and mechanisms from a variety of sources, and in some cases documents them in a style which will permit their use in many data different environments.

The term "method" is used in a very general sense to include any process, function, etc., that is relevant to QoS at any stage in the life-cycle of a system.

The criterion for reference to or inclusion of definitions or specifications of QoS methods and mechanisms in this Recommendation | Technical Report is that they are thought to be of potentially wider application than solely the environment for which they were originally developed, although still in a data context.

Clause 5 identifies sources of definitions of QoS characteristics and related information. Clauses 6, 7 and 8 discuss methods and mechanisms appropriate to the phases of QoS activity that are defined in the QoS Framework: clause 6 deals with the prediction phase, clause 7 with the establishment phase and clause 8 with the operational phase. Clause 9 describes methods for verification of system behaviour related to QoS. Clause 10 covers the relationships between this Recommendation | Technical Report and Recommendations, International Standards or Technical Reports that reference it.

This Recommendation | Technical Report contains detailed definitions of some QoS mechanisms. Some peer-to-peer QoS negotiation mechanisms are defined in 7.1.1. These involve two peer entities and in most cases also the provider of a communications service between them. Subclause 7.1.2 provides an initial specification of some QoS negotiation mechanisms for 1 × N multicast connections, based on those in 7.1.1. Subclause 7.1.3 discusses QoS negotiation mechanisms for M × N multicast, some of which can make use of those in 7.1.2. Subclause 8.2.1 defines some QoS management mechanisms to support time-critical applications.

This Recommendation | Technical Report does not include methods and mechanisms for security.

2 Normative references

The following Recommendations and International Standards contain provisions which, through reference in this text, constitute provisions of this Recommendation | Technical Report. 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 | Technical Report 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.641 (1997) | ISO/IEC 13236:1998, *Information technology – Quality of Service: Framework*.
- ITU-T Recommendation X.902 (1995) | ISO/IEC 10746-2:1996, *Information technology – Open distributed processing – Reference Model: Foundations.*

3 Definitions

For the purposes of this Recommendation | Technical Report, the following definitions apply.

3.1 QoS Framework definitions

This Recommendation | Technical Report uses the following terms drawn from the Quality of Service Framework, ITU-T Rec. X.641 | ISO/IEC 13236:

- compulsory (level of agreement);
- connection-wide (negotiation);
- controlled highest quality;
- establishment phase;
- guaranteed (level of agreement);
- highest quality attainable;
- lowest quality acceptable;
- operating target;
- operational phase;
- prediction phase;
- QoS alert;
- QoS characteristic;
- QoS enquiry;
- QoS filter;
- QoS maintenance;
- QoS measure;
- QoS mechanism;
- QoS monitoring;
- QoS negotiation;
- QoS parameter;
- QoS threshold;
- QoS verification;
- receiver-selected (negotiation).

3.2 ODP definitions

This Recommendation | Technical Report uses the following term drawn from the Reference Model of Open Distributed Processing: Foundations, 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.

4 Abbreviations

For the purposes of this Recommendation | Technical Report, the following abbreviations apply:

AGI	Active Group Identity
CHQ	Controlled Highest Quality
CORBA	Common Object Request Broker Architecture
CW	Connection-Wide (negotiation)
FDT	Formal Description Technique
HQA	Highest Quality Attainable
LQA	Lowest Quality Acceptable
ODP	Open Distributed Processing
OMG	Object Management Group
PCO&M	Point of Control, Observation and Measurement
PDU	Protocol Data Unit
QoS	Quality of Service
RFC	Request for Comment
RSVP	Resource Reservation (setup) Protocol
SUT	Service under Test
TCNM	Network Management for Time Critical Communications Systems

5 References to QoS from Recommendations, International Standards and other specifications

This clause identifies sources of definitions of QoS characteristics that have been developed by standards bodies and other organizations as an aid to those wishing to design QoS methods and mechanisms, together with sources of associated definitions and tutorial or other related information.

NOTE – ITU-T and ISO maintain catalogues of their Recommendations and International Standards, together with other useful information, on their World Wide Web servers at http://www.itu.int/ and http://www.iso.ch/ respectively. Readers are encouraged to consult these in order to ensure that they use the most up-to-date references.

5.1 QoS in collaborative ITU-T Recommendations and ISO/IEC International Standards

Standardization for various kinds of data transmission networks is undertaken within ITU-T and within ISO and IEC in ISO/IEC JTC 1. Frequently, the results of this effort are needed by both ITU-T and ISO/IEC and the work is therefore developed on a collaborative basis leading to identical or technically aligned ITU-T Recommendations and ISO/IEC International Standards. Both OSI and non-OSI environments are addressed, as are peer-to-peer and multipeer associations.

5.1.1 ITU-T Recommendations and ISO/IEC International Standards that reference QoS for the lower layers

The Recommendations and International Standards involved in this area cover service definitions, attachment standards and protocol specifications for a variety of technologies including:

- circuit-switched networks (both analogue and digital);
- packet-switched networks (e.g. those conforming to the ITU-T Rec. X.25);
- local and metropolitan area networks following the ISO/IEC 8802 family of standards;
- frame relay and broadband ISDN networks;
- simple peer-to-peer and multipeer data-links.

Many Recommendations | International Standards in this area make reference to various aspects of QoS. Table 5-1 lists the Recommendations | International Standards and shows which QoS characteristics they define or use. Table 5-2 provides a reverse index to the source documents: it lists the QoS characteristics or related QoS parameters that have been defined in the Recommendations | International Standards and identifies the documents in which they can be found. In addition, there is a *where-used* table in 7.1.1.2 (Table 7-1), which identifies the Recommendations | International Standards that utilize the QoS mechanisms defined in that subclause.

Table 5-1 – QoS characteristics and parameters in joint ITU-T and ISO/IEC lower layer standards

Recommendation | International Standard	Characteristic or parameter
ITU-T Rec. X.25 and ISO/IEC 8208 (X.25 Packet Layer Protocol)	throughput class transit delay selection and indication minimum throughput class end-to-end transit delay priority protection
ITU-T Rec. X.213 | ISO/IEC 8348 (Connection-mode network service definition)	throughput transit delay priority protection
ITU-T Rec. X.213 | ISO/IEC 8348 (Connectionless-mode network service definition)	transit delay cost determinants
ITU-T Rec. X.223 | ISO/IEC 8878 (Use of X.25 to provide the connection-mode network service)	throughput transit delay priority
ITU-T Rec. X.233 | ISO/IEC 8473-1 (Protocol for providing the connectionless-mode network service)	sequencing vs. transit delay transit delay vs. cost residual error probability vs. transit delay residual error probability vs. cost.
CCITT Rec. X.612 | ISO/IEC 9574 (Connection-mode network service by packet-mode terminal connected to an ISDN)	throughput transit delay
ITU-T Rec. X.622 | ISO/IEC 8473-3 (Connectionless-mode network protocol over X.25)	priority transit delay and throughput
ITU-T Rec. X.214 | ISO/IEC 8072 (Transport service definition)	establishment delay establishment failure probability throughput transit delay residual error rate transfer failure probability release delay release failure probability protection priority resilience
ITU-T Rec. X.224 | ISO/IEC 8073 (Connection-mode transport protocol)	throughput residual error rate transit delay priority
ITU-T Rec. X.234 | ISO/IEC 8602 (Connectionless-mode transport protocol)	QoS parameter defined by connectionless-mode transport service
ITU-T Rec. X.605 | ISO/IEC 13252 (Enhanced communications transport service definition)	throughput transit delay transit delay jitter corrupted data unit error rate lost data unit error rate ordering protection precedence

Table 5-2 – Index to sources of definitions of QoS characteristics and parameters

Characteristic or parameter	Recommendation | International Standard
transit delay	ITU-T Rec. X.25 and ISO/IEC 8208 ITU-T Rec. X.213 | ISO/IEC 8348 – CO ITU-T Rec. X.213 | ISO/IEC 8348 – CL ITU-T Rec. X.214 | ISO/IEC 8072 ITU-T Rec. X.223 | ISO/IEC 8878 ITU-T Rec. X.224 | ISO/IEC 8073 ITU-T Rec. X.233 | ISO/IEC 8473-1 ITU-T Rec. X.605 | ISO/IEC 13252 CCITT Rec. X.612 | ISO/IEC 9574 ITU-T Rec. X.622 | ISO/IEC 8473-3
transit delay jitter	ITU-T Rec. X.605 | ISO/IEC 13252
establishment delay	ITU-T Rec. X.214 | ISO/IEC 8072
release delay	ITU-T Rec. X.214 | ISO/IEC 8072
throughput	ITU-T Rec. X.25 and ISO/IEC 8208 ITU-T Rec. X.213 | ISO/IEC 8348 – CO ITU-T Rec. X.214 | ISO/IEC 8072 ITU-T Rec. X.223 | ISO/IEC 8878 ITU-T Rec. X.224 | ISO/IEC 8073 ITU-T Rec. X.605 | ISO/IEC 13252 CCITT Rec. X.612 | ISO/IEC 9574 ITU-T Rec. X.622 | ISO/IEC 8473-3
protection	ITU-T Rec. X.25 and ISO/IEC 8208 ITU-T Rec. X.213 | ISO/IEC 8348 – CO ITU-T Rec. X.214 | ISO/IEC 8072 ITU-T Rec. X.605 | ISO/IEC 13252
residual error rate	ITU-T Rec. X.214 | ISO/IEC 8072 ITU-T Rec. X.224 | ISO/IEC 8073 ITU-T Rec. X.233 | ISO/IEC 8473-1 ITU-T Rec. X.605 | ISO/IEC 13252
establishment failure probability	ITU-T Rec. X.214 | ISO/IEC 8072
transfer failure probability	ITU-T Rec. X.214 | ISO/IEC 8072
release failure probability	ITU-T Rec. X.214 | ISO/IEC 8072
resilience	ITU-T Rec. X.214 | ISO/IEC 8072
priority/precedence	ITU-T Rec. X.25 and ISO/IEC 8208 ITU-T Rec. X.213 | ISO/IEC 8348 – CO ITU-T Rec. X.214 | ISO/IEC 8072 ITU-T Rec. X.223 | ISO/IEC 8878 ITU-T Rec. X.224 | ISO/IEC 8073 ITU-T Rec. X.605 | ISO/IEC 13252 ITU-T Rec. X.622 | ISO/IEC 8473-3
ordering	ITU-T Rec. X.605 | ISO/IEC 13252
sequencing	ITU-T Rec. X.233 | ISO/IEC 8473-1
cost determinants	ITU-T Rec. X.213 | ISO/IEC 8348 – CL ITU-T Rec. X.233 | ISO/IEC 8473-1

The following Recommendations | International Standards cover service definitions:

- ITU-T Recommendation X.213 (1995) | ISO/IEC 8348:1996, *Information technology – Open Systems Interconnection – Network service definition.*
- ITU-T Recommendation X.214 (1995) | ISO/IEC 8072:1996, *Information technology – Open Systems Interconnection – Transport service definition.*
- ITU-T Recommendation X.605 (1998) | ISO/IEC 13252:1999, *Information technology – Enhanced communications transport service definition.*

The following Recommendations | International Standards cover generalized protocol specifications:

- ITU-T Recommendation X.25 (1996), *Interface between Data Terminal Equipment (DTE) and Data Circuit-terminating Equipment (DCE) for terminals operating in the packet mode and connected to public data networks by dedicated circuit.*
- ISO/IEC 8208:1995, Information technology – Data Communications – X.25 Packet Layer Protocol for Data Terminal Equipment.
- ITU-T Recommendation X.224 (1995) | ISO/IEC 8073:1997, *Information technology – Open Systems Interconnection – Protocol for providing the connection-mode transport service.*
- ITU-T Recommendation X.233 (1997) | ISO/IEC 8473-1:1998, *Information technology – Protocol for providing the connectionless-mode network service: Protocol specification.*
- ITU-T Recommendation X.234 (1994) | ISO/IEC 8602:1995, *Information technology – Protocol for providing the OSI connectionless-mode transport service.*

The following Recommendations | International Standards cover protocol specifications for specific technologies:

- ITU-T Recommendation X.223 (1993), *Use of X.25 to provide the OSI connection-mode network service for ITU-T applications.*

  ISO/IEC 8878:1992, *Information technology – Telecommunications and information exchange between systems – Use of X.25 to provide the OSI Connection-mode network Service.*
- CCITT Recommendation X.612 (1992) | ISO/IEC 9574:1992, *Information technology – Provision of the OSI connection-mode network service by packet-mode terminal equipment connected to an Integrated Services Digital Network (ISDN).*
- CCITT Recommendation X.613 (1992) | ISO/IEC 10588:1993, *Information technology – Use of X.25 Packet Layer Protocol in conjunction with X.21/X.21 bis to provide the OSI connection-mode Network service.*
- CCITT Recommendation X.614 (1992) | ISO/IEC 10732:1993, *Information technology – Use of X.25 Packet Layer Protocol to provide the OSI connection-mode Network service over the telephone network.*
- ITU-T Recommendation X.622 (1994) | ISO/IEC 8473-3:1995, *Information technology – Protocol for providing the connectionless-mode Network service: Provision of the underlying service by an X.25 subnetwork.*
- ITU-T Recommendation X.623 (1994) | ISO/IEC 8473-4:1995, *Information technology – Protocol for providing the connectionless-mode network service: Provision of the underlying service by a subnetwork that provides the OSI Data Link service.*
- ITU-T Recommendation X.625 (1996) | ISO/IEC 8473-5:1997, *Information technology – Protocol for providing the connectionless-mode Network service: Provision of the underlying service by ISDN circuit-switched B-channels.*

5.1.2 ITU-T Recommendations and ISO/IEC International Standards that reference QoS for the upper layers

The Recommendations | International Standards for OSI higher-layer service definitions and protocol specifications that make reference to QoS are:

- ITU-T Recommendation X.215 (1995) | ISO/IEC 8326:1996, *Information technology – Open Systems Interconnection – Session service definition.*
- ITU-T Recommendation X.216 (1994) | ISO/IEC 8822:1994, *Information technology – Open Systems Interconnection – Presentation service definition.*

- ITU-T Recommendation X.217 (1995) | ISO/IEC 8649:1996, *Information technology – Open Systems Interconnection – Service definition for the Association Control Service Element.*
- ITU-T Recommendation X.225 (1995) | ISO/IEC 8327-1:1996, *Information technology – Open Systems Interconnection – Connection-oriented session protocol: Protocol specification.*
- ITU-T Recommendation X.226 (1994) | ISO/IEC 8823-1:1994, *Information technology – Open Systems Interconnection – Connection-oriented presentation protocol: Protocol specification.*
- ITU-T Recommendation X.227 (1995) | ISO/IEC 8650-1:1996, *Information technology – Open Systems Interconnection – Connection-oriented protocol for the Association Control Service Element: Protocol specification.*
- ITU-T Recommendation X.235 (1995) | ISO/IEC 9548-1:1996, *Information technology – Open Systems Interconnection – Connectionless session protocol: Protocol specification.*
- ITU-T Recommendation X.236 (1995) | ISO/IEC 9576-1:1995, *Information technology – Open Systems Interconnection – Connectionless presentation protocol: Protocol specification.*
- ITU-T Recommendation X.237 (1995) | ISO/IEC 10035-1:1995, *Information technology – Open Systems Interconnection – Connectionless protocol for the Association Control Service Element: Protocol specification*.

The Recommendations | International Standards for Message Handling Systems (MHS) that make reference to QoS are:

- ITU-T Recommendation X.400 series | ISO/IEC 10021 (all parts), *Information technology – Message Handling Systems (MHS).*

The following Recommendations | International Standards for OSI systems management are specifications that support QoS management:

- ITU-T Recommendation X.701 (1997) | ISO/IEC 10040:1998, *Information technology – Open Systems Interconnection – Systems management overview.*
- ITU-T Recommendation X.710 (1997) | ISO/IEC 9595:1998, *Information technology – Open Systems Interconnection – Common management information service.*
- ITU-T Recommendation X.711 (1997) | ISO/IEC 9596-1:1998, *Information technology – Open Systems Interconnection – Common management information protocol: Specification.*
- ITU-T Recommendation X.72x series | ISO/IEC 10165 (all parts), *Information technology – Open Systems Interconnection – Structure of management information.*
- ITU-T Recommendations X.730 to X.753 | ISO/IEC 10164 (all parts), *Information technology – Open Systems Interconnection – Systems management.*

5.1.3 ITU-T Recommendations and ISO/IEC International Standards that reference QoS for Open Distributed Processing

The following Recommendations | International Standards for ODP make reference to QoS:

- ITU-T Recommendations X.901 to X.904 | ISO/IEC 10746 (Parts 1 to 4), *Information technology – Open distributed processing – Reference Model.*
- ITU-T Recommendation X.950 (1997) | ISO/IEC 13235-1:1998, *Information technology – Open distributed processing – Trading function: Specification.*

Work on QoS in ODP is under way to produce Recommendations | International Standards including:

- a new part of the Reference Model of Open Distributed Processing; ITU-T Rec. X.90x | ISO/IEC 10746-x;
- any necessary amendments to the other parts of the Reference Model of Open Distributed Processing, ITU-T Recs. X.901 to X.904 | ISO/IEC 10746 (Parts 1 to 4), to reference or summarize the new part and provide alignment;
- other stand-alone Recommendations | International Standards for QoS in ODP as necessary.

It is intended to develop these specifications in collaboration with the Object Management Group (OMG), which is extending its activities to include the specification of QoS in CORBA-based systems. It is expected that it will be possible to agree upon a significant amount of common text.

5.2 QoS in ISO/IEC International Standardized Profiles

International Standardized Profiles (ISPs) that reference the OSI protocol specifications listed above may place constraints on the treatment of QoS. The taxonomy of profiles is provided in ISO/IEC TR 10000, Framework and Taxonomy of Profiles.

5.3 QoS in ISO TC 184 standards

This subclause identifies ISO TC 184 documents that contain information on current areas of TC 184 work relating to QoS:

– ISO TR 12178:1994, *Industrial automation – Time-critical communications architectures – User requirements.*

References to further TC 184 standards dealing with QoS can be found in ISO TR 12178.

5.4 QoS in ITU-T Recommendations

This subclause identifies some ITU-T Recommendations that contain definitions of QoS characteristics and/or related information.

5.4.1 QoS in G-series Recommendations – Transmission systems and media, digital systems and networks

– ITU-T Recommendation G.826 (1996), *Error performance parameters and objectives for international, constant bit rate digital paths at or above the primary rate.*

– ITU-T Recommendation G.827 (1996), *Availability parameters and objectives for path elements of international constant bit-rate digital paths at or above the primary rate.*

5.4.2 QoS in I-series Recommendations – Integrated Services Digital Networks

– ITU-T Recommendation I.350 (1993), *General aspects of quality of service and network performance in digital networks, including ISDNs.*

– ITU-T Recommendation I.351 (1997), *Relationships among ISDN performance Recommendations.*

– ITU-T Recommendation I.352 (1993), *Network performance objectives for connection processing delays in an ISDN.*

– ITU-T Recommendation I.353 (1996), *Reference events for defining ISDN and B-ISDN performance parameters.*

– ITU-T Recommendation I.354 (1993), *Network performance objectives for packet-mode communication in an ISDN.*

– ITU-T Recommendation I.355 (1995), *ISDN 64 kbit/s connection type availability performance.*

– ITU-T Recommendation I.356 (1996), *B-ISDN ATM layer cell transfer performance.*

– ITU-T Recommendation I.357 (1996), *B-ISDN semi-permanent connection availability*.

5.4.3 QoS in X-series Recommendations – Data networks and open system communication

– CCITT Recommendation X.130 (1988), *Call processing delays in public data networks when providing international synchronous circuit-switched data services.*

– CCITT Recommendation X.131 (1988), *Call blocking in public data networks when providing international synchronous circuit-switched data services.*

– ITU-T Recommendation X.134 (1997), *Portion boundaries and packet layer reference events: Basis for defining packet-switched performance parameters.*

– ITU-T Recommendation X.135 (1997), *Speed of service (delay and throughput) performance values for public data networks when providing international packet-switched services.*

– ITU-T Recommendation X.136 (1997), *Accuracy and dependability performance values for public data networks when providing international packet-switched services.*

– ITU-T Recommendation X.137 (1997), *Availability performance values for public data networks when providing international packet-switched services.*

- ITU-T Recommendation X.138 (1997), *Measurement of performance values for public data networks when providing international packet-switched services.*
- ITU-T Recommendation X.139 (1997), *Echo, drop, generator and test DTEs for measurement of performance values in public data networks when providing international packet-switched services.*
- CCITT Recommendation X.140 (1992), *General quality of service parameters for communication via public data networks.*
- CCITT Recommendation X.141 (1988), *General principles for the detection and correction of errors in public data networks.*
- ITU-T Recommendation X.144 (1995), *User information transfer performance parameters for data networks providing international frame relay PVC service.*
- ITU-T Recommendation X.145 (1996), *Performance for data networks providing international frame relay SVC service.*
- ITU-T Recommendation X.146 (1998), *Performance objectives and quality of service classes applicable to frame relax.*

5.5 QoS in specifications produced by other organizations

5.5.1 Internet Engineering Task Force (IETF)

Considerable interest is now being shown in introducing management of QoS into the Internet. The following Internet Requests for Comment (RFCs) make reference to QoS or are relevant to it:

- RFC 1633 (June 1994), *Integrated Services in the Internet Architecture: An Overview.*
- RFC 1819 (August 1994), *Internet Stream Protocol Version 2 (ST2) Protocol Specification – Version ST2+.*
- RFC 1821 (August 1995), *Integration of Real-time Services in an IP-ATM Network Architecture.*
- RFC 1883 (December 1995), *Internet Protocol, Version 6 (IPv6) Specification.*
- RFC 1889 (January 1996), *RTP: A Transport Protocol for Real-Time Applications.*
- RFC 2205 (September 1997), *Resource ReSerVation Protocol (RSVP) – Version 1 Functional Specification.*
- RFC 2206 (September 1997), *RSVP Management Information Base using SMIv2.*
- RFC 2207 (September 1997), *RSVP Extensions for IPSEC Data Flows.*
- RFC 2208 (September 1997), *Resource ReSerVation Protocol (RSVP) – Version 1 Applicability Statement Some Guidelines on Deployment.*
- RFC 2209 (September 1997), *Resource ReSerVation Protocol (RSVP) – Version 1 Message Processing Rules.*
- RFC 2210 (September 1997), *The Use of RSVP with IETF Integrated Services.*
- RFC 2211 (September 1997), *Specification of the Controlled-Load Network Element Service.*
- RFC 2212 (September 1997), *Specification of Guaranteed Quality of Service.*
- RFC 2213 (September 1997), *Integrated Services Management Information Base using SMIv2.*
- RFC 2214 (September 1997), *Integrated Services Management Information Base Guaranteed Service Extensions using SMIv2.*
- RFC 2215 (September 1997), *General Characterization Parameters for Integrated Service Network Elements.*
- RFC 2216 (September 1997), *Network Element Service Specification Template*.

These and other RFCs are available from various directories, including the directory ftp://ds.internic.net/rfc. Individual RFCs can be accessed using ftp://ds.internic.net/rfc/rfc*nnnn*.txt, where *nnnn* is the RFC number.

Further information, including relevant Internet-Drafts, is available from the IETF Web pages at http://www.ietf.org/. More details on the various Internet Working Groups and the documents they have produced can be found at this site. The following Working Groups are of particular relevance to QoS management:

- Benchmarking Methodology, concerned with performance measurement:
  http://www.ietf.org/html.charters/bmwg-charter.html;

- Integrated Services, concerned with the transport of audio, video, real-time and classical data traffic within a single network infrastructure, and defining *best-effort*, *control load* and *guaranteed* services: http://www.ietf.org/html.charters/intserv-charter.html;
- Integrated Services over Specific Link Layers: http://www.ietf.org/html.charters/issll-charter.html;
- QoS Routing: http://www.ietf.org/html.charters/qosr-charter.html;
- Realtime Traffic Flow Measurement: http://www.ietf.org/html.charters/rtfm-charter.html;
- Resource Reservation Setup Protocol (RSVP): http://www.ietf.org/html.charters/rsvp-charter.html.

There is also a Web page for RSVP at http://www.isi.edu/div7/rsvp/ietf.html.

5.6 Research on QoS management

QoS management is now an important research topic for many universities, institutes, consortia and industrial organizations. The number of papers published on the subject is growing each year, to an extent that makes it impracticable for this Recommendation | Technical Report to give individual references. However, the proceedings of the annual *IFIP International Workshop on Quality of Service* (IWQoS) may be a useful starting point for those interested in research in QoS: the fifth IWQoS was held in May 1997. IWQoS has a Web page at http://www.ctr.columbia.edu/iwqos/.

6 Methods and mechanisms for the prediction phase

As defined in the QoS Framework, the QoS prediction phase includes the following activities:

- 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 (e.g. 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;
- checking that requests will not conflict with admission control policies.

Typically, such mechanisms are implemented by local or proprietary means, and no standards or publicly available specifications have been identified containing relevant specifications. Standard OSI or Internet Management can be used to support communications where needed as part of prediction phase activities.

7 Methods and mechanisms for the establishment phase

This clause identifies methods and mechanisms for the QoS establishment phase, as defined in the QoS Framework.

Mechanisms for the establishment phase include:

- methods of reaching QoS agreements, including negotiation mechanisms;
- resource allocation mechanisms;
- initialization mechanisms.

7.1 Methods of reaching QoS agreements

QoS agreements can be reached by a variety of means, including:

- administration, e.g. as part of a subscription process;
- imposition by one of the parties to the interaction;
- negotiation;
- management or security policy.

The remainder of this subclause discusses QoS negotiation in the context of communications protocols, under the headings:

- negotiating QoS in peer-to-peer communications;
- negotiating QoS in $1 \times N$ multicast;
- negotiating QoS in $M \times N$ multicast.

7.1.1 Negotiating QoS in peer-to-peer communications

The concepts of negotiation of QoS between two peer entities are discussed in the QoS Framework, ITU-T Rec. X.641 | ISO/IEC 13236, 7.3 and 8.3.2. Such negotiation may involve just the two peer entities alone, or may include the provider of a communications service between them. Many negotiation mechanisms have been developed for the lower layer communications protocols standardized jointly by ITU-T and ISO/IEC.

Generalizing those mechanisms, this clause defines two basic three-party negotiation mechanisms involving two users and a provider that may be used to reach QoS agreements of the types defined in clause 7 of the QoS Framework. The first uses a single parameter, and permits negotiation down from proposed maximum or desirable QoS levels. The second permits the parties to specify ranges in which they are capable of operating, and enables them to agree upon a limit, an operating target or a threshold within those ranges.

Multiple instances of the mechanisms defined in this clause may be operated in order to negotiate combinations of operating target, limits and/or thresholds. Thus, for example, agreements on high and low limits can be reached by operating two single-parameter negotiation mechanisms simultaneously.

NOTE – However, in complex cases it may be better to define new combined mechanisms to achieve the same result more efficiently.

Although each mechanism has a defined order of operation, with an initiating user proposing an initial value or values, which are then modified by the other parties, there is a degree of symmetry in that the outcome of the negotiation should be a value that is acceptable to all. However, there is an asymmetry in that some parties can exercise their choice of value within the acceptable region.

The two mechanisms are defined in 7.1.1.1, and then 7.1.1.2 illustrates their use for reaching agreements of the kinds defined in the QoS Framework. Subclause 7.1.1.3 shows how the use of these mechanisms to negotiate various QoS characteristics has been specified in ITU-T Recommendations | ISO/IEC International Standards.

It should be noted that, in this clause, mechanisms are described which include the negotiation of both upper and lower limits. Although these are included as examples of possible mechanisms, it is recognized that negotiation mechanisms which involve a single limit (either high or low) are likely, in general, to be found in real systems and networks.

7.1.1.1 Basic three-party negotiation mechanisms

In the following, the terms increase, high value, better value and upper bound are all to be understood as in the direction of higher quality, and the terms decrease, low value, worse value and lower bound are to be understood as in the direction of lower quality. High quality values may be either high numerically (as in the case of throughput) or low (as in the case of transit delay).

In the mechanism descriptions which follow, actions undertaken by a provider are described. These commonly involve selection by a provider, during the negotiation phase, of a new value for a QoS parameter and are often described in the following manner: the provider may select a new value P' which is not better than the initiator-proposed value, i.e. such that $P' \leq P$. These inequalities are chosen so that the negotiation mechanisms converge and terminate and the values P' and P represent the values actually exchanged as part of the negotiation mechanism. This does not preclude a provider from operating internally at a higher quality than P, but this would not be signalled in the negotiation mechanism. It is recognized that many networks, for example, operate on a discrete number of settings for a particular QoS characteristic and that in practice, therefore, a higher quality than that requested for a particular characteristic may well be provided.

The following definitions cover cases of normal operation. Other behaviour may occur in cases of equipment failure or sudden overload conditions.

a) Single-parameter negotiation

1) The initiating user supplies a proposed value P to the provider.

2) The provider may refuse the request. If the provider does not refuse the request, it may select a new proposed value P' which is not better than the initiator-proposed value, i.e. such that $P' \leq P$. The provider supplies the (possibly revised) proposed value to the responding user.

3) The responding user may refuse the request. If the responding user does not refuse the request, it may select a new value V which is not better than the provider proposed value, i.e. such that $V \leq P' \leq P$.

4) The provider shall leave the selected value V unchanged.

5) The selected value V is returned to the initiating user. It is the "agreed" value.

b) Bounded negotiation

1) The initiating user specifies a desired operating range by supplying to the provider a lower bound L and an upper bound U, where L ≤ U. (Where an LQA limit is being negotiated, L is the proposed LQA value. Where an operating target is being negotiated, U is the proposed target value. Where a CHQ limit is being negotiated, U is the proposed CHQ value.)

2) The provider may refuse the request if it knows it cannot be met, i.e. if it cannot support at least the lower bound value L. If the provider does not refuse the request, but cannot operate over the full range proposed by the initiating user, it may determine a new reduced value U′ for the upper bound: this reduced value may not be worse than the lower bound. Thus L < U′ ≤ U. (It is possible that the provider may choose to operate internally at a higher quality, but it does not signal this fact to the responding user.)

   The provider may not alter the lower bound L. The new upper bound U′ and the lower bound L are supplied to the responding user.

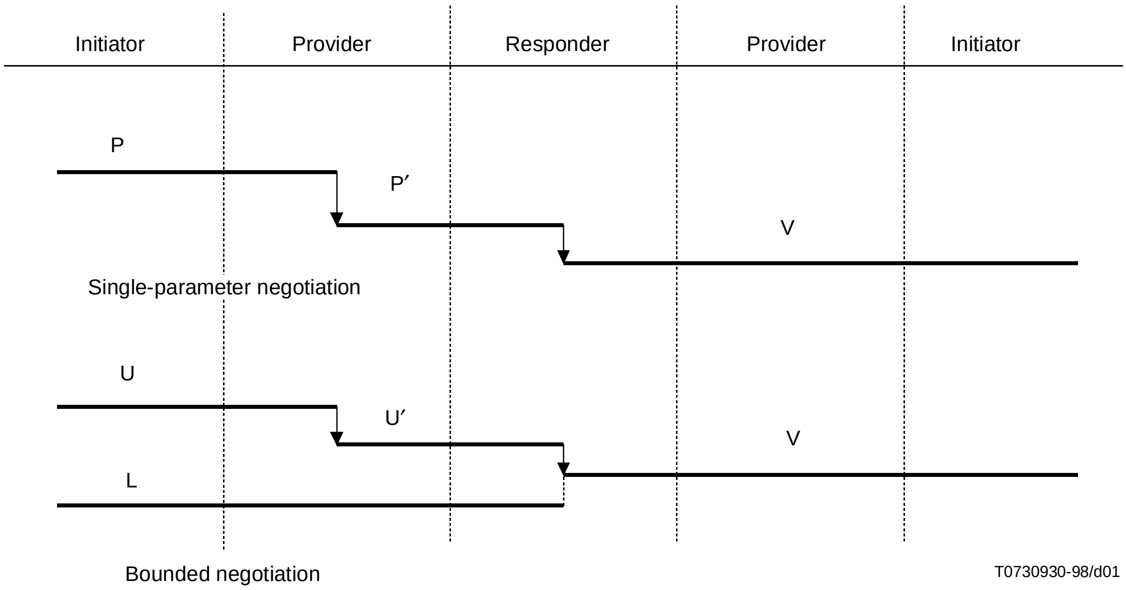
3) The responding user may refuse the request. If it accepts, it may select any value V in the range between the lower and upper bounds supplied. Thus L ≤ V ≤ U′. The selected value is returned to the provider.

4) The provider shall leave the selected value V unchanged.

5) The selected value V is returned to the initiating user. It is the "agreed" value.

The two mechanisms are illustrated in Figure 7-1.

The mechanisms may also be operated with some restrictions placed on the behaviour of one or more of the parties. For example, bilateral negotiation mechanisms correspond to restricted versions of the above where the provider is not permitted to modify any of the values it receives but must pass them unchanged to the other user. Also, thresholds may often be negotiated simply between one user and the provider, where that user wishes to be informed when QoS reaches a certain level but the provider's capability also has to be taken into account.



NOTE – The provider and responding user may refuse to accept the proposed value(s) and thus abort the negotiation.

Figure 7-1 – Three-party negotiation

7.1.1.2 Application of the negotiation mechanisms

The mechanisms defined above can be used to negotiate QoS agreements on individual values relating to QoS characteristics. Specifically, such a QoS agreement includes:

- the QoS characteristic to which it relates;
- where applicable, whether it relates to a given direction of transfer, or to both directions equally;
- the type of value negotiated, which may be an operating target, a limit or a threshold: these are defined in the QoS Framework (see 7.3.2);
- the level of agreement: these are defined in the QoS Framework (see 7.3.2.4 in ITU-T Rec. X.641 | ISO/IEC 13236), and may be:
  - best-efforts, for operating targets or thresholds;
  - compulsory, for limits only;
  - guaranteed, for limits only.

In an environment where many different types of QoS agreement may be negotiated, it may be necessary to use several service or protocol QoS parameters to convey the precise intention of a negotiation step.

Single-parameter negotiation is most appropriate to the negotiation of:

- a Highest Quality Attainable (HQA) operating target, where the desire is to operate at the highest level acceptable to all parties, with best-effort semantics;
- a Controlled Highest Quality (CHQ) limit, where the desire is to establish an upper limit on QoS, and the consequent agreement may have best-effort, compulsory or guaranteed semantics;
- a high threshold value.

Bounded negotiation is most appropriate to negotiation of:

- a Highest Quality Attainable (HQA) operating target, where the desire is to operate at the highest level acceptable to all parties, with best-effort semantics, but with a lower bound on acceptable QoS;
- a Controlled Highest Quality (CHQ) limit, where the desire is to establish an upper limit on QoS but with a lower bound on that limit, and the consequent agreement may have best-effort, compulsory or guaranteed semantics;
- a Lowest Quality Acceptable (LQA) limit, where the desire is to establish a level below which the QoS should not fall, and the consequent agreement may have best-effort, compulsory or guaranteed semantics;
- a high or low threshold value.

Where bounded negotiation is used, there is the further question of where in the available range the responding user should select the final value [in step 3) of the definition of bounded negotiation in 7.1.1.1]. Typically, high final values would be appropriate when negotiating HQA, CHQ or high threshold values; and low values would be appropriate when negotiating LQA or low threshold values.

7.1.1.3 Usage of mechanisms mapped to lower layer ITU-T Recommendations | ISO/IEC International Standards

Table 7-1 below identifies, for those standards which specify a particular mechanism for use in conjunction with a given QoS characteristic, which of the mechanisms defined in this clause is utilized. It also identifies cases where the standards state that a characteristic is not negotiated. If there is no specific entry in this table for a given standard, then that standard does not specify a particular mechanism, and relies upon a characteristic by characteristic mapping to take place from the upper service or protocol to its lower counterpart.

NOTE 1 – In this *where-used* table, only an abbreviated form of title is given. For the full title, see the lists of Recommendations | International Standards in 5.1.1.

NOTE 2 – Although Table 7-1 includes text extracted from, or a summary of, the Recommendations | International Standards or Recommendations | Technical Reports referenced, in the event of any conflict between the statements in this guide and those in the documents referenced, then the requirements as stated in the documents referenced shall take precedence.

NOTE 3 – Owing to their complexity, the mechanisms developed for the Enhanced Communications Transport Service and Protocol are not included in Table 7-1. A brief discussion and references are given in 7.1.3.

Table 7-1 – Usage of mechanisms mapped to lower layer ITU-Recommendations | ISO/IEC International Standards

Recommendation | International Standard	Characteristic or parameter	Mechanism used	Notes
ITU-T Rec. X.25 and ISO/IEC 8208 (X.25 Packet Layer Protocol)	throughput class	Single-parameter negotiation of an HQA operating target, with best-efforts semantics.	
	transit delay selection and indication	1) The initiating user supplies a proposed value. 2) The provider honours the request when it can do so. 3) The responding user and initiating user are informed of the value selected. It may be less than, equal to, or greater than the proposed value.	
	minimum throughput class	1) The initiating user supplies a proposed value. 2) The provider may refuse the request. If it does not do so it forwards the proposed value to the responding user. 3) The responding may refuse the request, or accept the proposed value.	
	end-to-end transit delay	At the choice of the initiator, either bounded negotiation or single-parameter negotiation of an HQA operating target value, with best efforts semantics.	
	priority	Bounded negotiation of an HQA operating target value, with best-efforts semantics.	
	protection	Bounded negotiation of an HQA operating target value, with best-efforts semantics.	
ITU-T Rec. X.213 | ISO/IEC 8348 (Connection-mode network service definition)	throughput	At the choice of the initiator, either bounded negotiation or single-parameter negotiation of an HQA operating target value, with best efforts semantics.	
	transit delay	Bounded negotiation or single-parameter negotiation of an HQA operating target value, with best efforts semantics.	1
	priority	At the choice of the initiator, either bounded negotiation or single-parameter negotiation of an HQA operating target value, with best efforts semantics.	
	protection	Not negotiated – A local matter controlled according to the security policy in force. See ITU-T Rec. X.802 | ISO/IEC TR 13594 (Information technology – Lower Layers security model).	
	transit delay cost determinants	Not a negotiation mechanism but based on the specific characteristics of the facilities which can be expected to be made available by the provider.	
ITU-T Rec. X.214 | ISO/IEC 8072 (Transport service definition)	all QoS parameters except protection	Single-parameter negotiation of an HQA operating target, with best-efforts semantics.	
	protection	Not negotiated – A local matter controlled according to the security policy in force. See ITU-T Rec. X.802 | ISO/IEC TR 13594 (Information technology – Lower Layers security model).	
ITU-T Rec. X.223 | ISO/IEC 8878 (Use of X.25 to provide the connection-mode network service)	throughput transit delay priority	At the choice of the initiator, either bounded negotiation or single-parameter negotiation of an HQA operating target value, with best efforts semantics.	
ITU-T Rec. X.224 | ISO/IEC 8073 (Connection-mode transport protocol)	throughput residual error rate transit delay	Bounded negotiation of an HQA operating target, with best-efforts semantics.	
	priority	Single-parameter negotiation of an HQA operating target, with best-efforts semantics.	

Table 7-1 *(concluded)*

Recommendation | International Standard	Characteristic or parameter	Mechanism used	Notes
ITU-T Rec. X.233 | ISO/IEC 8473-1 (Protocol for providing the connectionless-mode network service)	sequencing vs. transit delay transit delay vs. cost residual error probability vs. transit delay residual error probability vs. cost	No negotiation – Decision taken on each individual data unit. Network entities in intermediate systems may, but are not required to, make use of this information as an aid in selecting a route when more than one route satisfying other routing criteria is available and the available routes are known to differ with respect to QoS. When this information is used, routing decisions should favour the QoS usage indicated by the user.	
ITU-T Rec. X.234 | ISO/IEC 8602 (Connectionless-mode transport protocol)	QoS parameter defined by connectionless-mode Transport service	For underlying connectionless-mode network service: QoS parameter derived from the *a priori* knowledge by the user of the QoS available on the association. For underlying connection-mode network service: QoS parameter derived from knowledge by the user of the QoS available from the network connection.	
CCITT Rec. X.612 | ISO/IEC 9574 (Connection-mode network service by packet-mode terminal connected to an ISDN)	throughput transit delay	Based on *a priori* knowledge – This knowledge may be modified by ITU-T Rec. Q.931 (ISDN user-network interface layer 3 specification for basic call control).	
ITU-T Rec. X.622 | ISO/IEC 8473-3 (Connectionless-mode network protocol over X.25)	priority	No negotiation mechanism	
	transit delay and throughput	Uses ITU-T Rec. X.25 and ISO/IEC 8208 protocol negotiation mechanisms.	2
NOTE 1 – For the transit delay case, a restricted form of the bounded negotiation mechanism is employed. In step 3), if the responding user accepts the request, then the supplied upper value (which corresponds to low transit delay) is to be used. It is not even conveyed in step 4), since the provider already knows the value that is to be used if the network-connection is set up. NOTE 2 – See the entry for ITU-T Rec. X.25 and ISO/IEC 8208.			

7.1.2 Negotiating QoS in 1 × N multicast

A 1 × N multicast connection is a special case of multipeer communications, namely a connection with one sender and N receivers. In establishing a 1 × N connection, there is in general an interaction between the negotiation of QoS and the selection of the participants in the connection. For example, a QoS requirement imposed by the sender may be beyond the capacity of the provider in some region, and hence some receivers may be excluded from participation in the connection. Receivers may also choose not to participate for reasons unrelated to QoS. The general treatment of 1 × N connection establishment is outside the scope of this Recommendation | Technical Report; this subclause covers only the negotiation of QoS.

NOTE 1 – Mechanisms for group selection for Enhanced Communications Transport Service (ECTS) are defined in ITU-T Rec. X.605 | ISO/IEC 13252.

NOTE 2 – The terminology of "sender" and "receiver" in 1 × N multicast is common, and is based on a typical application, namely broadcast transmission. However, this should not be taken to imply that the receivers cannot send data to the sender. In general, the possible transmission modes in 1 × N multicast are: sender to receivers multicast, sender to a single receiver unicast, and receiver to sender unicast. The term "full duplex 1 × N multicast" is used in cases where all these possibilities are exploited.

When negotiating QoS for 1 × N multicast connections, it is necessary to choose between two types of QoS negotiation mechanisms:

– "connection-wide" QoS negotiation mechanisms, which negotiate the same value of a QoS characteristic for the sender, the service provider and all receivers; and

– "receiver-selected" QoS negotiation mechanisms, which negotiate separate values of a QoS characteristic for each receiver, representing an agreement between the sender, the service provider and that particular receiver.

Different types of negotiation mechanism can be chosen for different characteristics in the establishment of a single 1 × N multicast connection. The choice for any particular characteristic may depend on the application: it is not necessarily an inherent property of the characteristic itself.

Connection-wide negotiation mechanisms may be applied to QoS characteristics:

- which by their very definition apply to the 1 × N multicast connection as a whole, such as protection (for confidentiality); or
- for which the value perceived by the receiver is dependent upon the behaviour of the sender or of the service provider as a whole, or for application reasons must be the same for all receivers. Examples could be throughput, in cases where data may not be lost; or transit delay, in cases where it is required to ensure synchronization across multiple receivers.

Receiver-selected negotiation mechanisms may be applied to those QoS characteristics for which there exists no application requirement to agree on a connection-wide value. Examples could be transit delay; or throughput, if loss of data can be tolerated.

Receiver-selected negotiation for a 1 × N multicast connection is performed for each of the N receivers independently. Hence the mechanisms defined in 7.1.1.1 for basic three-party QoS negotiation can be used for each such negotiation without change.

For connection-wide negotiation mechanisms, however, it is necessary to extend the mechanisms defined in 7.1.1.1 to ensure that a single QoS agreement is negotiated that is common to all receivers, and that the negotiated value is consistent with any constraints identified during the negotiation process, such as a limit on provider capacity local to one of the receivers. The necessary extensions are defined in 7.1.2.1 below, which also identifies how they may be applied to reaching agreements of the kinds defined in the QoS Framework.

Finally, 7.1.2.2 discusses the use of filters in 1 × N multicast.

7.1.2.1 Mechanisms for connection-wide QoS negotiation in 1 × N connections

When a QoS characteristic is to be negotiated connection-wide, it is necessary to inform all the participants of the result of the negotiation in a "three-way handshake". This adds a further step to the mechanisms of 7.1.1.1.

Other modifications to the mechanisms are required in order to determine an agreed QoS value that is consistent with all the requirements and constraints expressed by the sender, provider and the N receivers. In some cases, it will be possible to achieve a consistent value only by the exclusion of some receivers; when that happens, it may also be possible to achieve consistent values in different ways, with the exclusion of different sets of receivers. The choice of which receivers to exclude in such cases is outside the scope of this Recommendation | Technical Report.

NOTE 1 – Where it is necessary to negotiate multiple limits, or an operating target together with one or more limits, it is possible to operate multiple instances of the mechanisms defined in this subclause. However, it may be better to define other mechanisms to deal with these cases more efficiently.

As in 7.1.1.1, the terms increase and decrease are to be understood as meaning changes in the direction of improved or degraded quality respectively. In 1 × N multicast, it is the multicast sender that is the initiating user, and the receivers are the responding users.

The procedures defined below are those for normal operation. Other behaviour may occur in cases of equipment failure or sudden overload conditions.

NOTE 2 –The definitions of the negotiation mechanisms that follow treat the "service-provider" as a single participant, even though in practice it may be composed of a number of communications providers.

Five mechanisms are defined, as follows:

- "single-parameter" negotiation, which is negotiation down from upper bounds provided by the parties in succession, with no lower bounds imposed;
- bounded negotiation of a low limit or a low threshold;
- bounded negotiation of a high limit or a high threshold;
- bounded negotiation of an operating target;
- combined negotiation of upper and lower limits.

The term "bounded" is used to characterize negotiation mechanisms in which bounds are placed on how far values may be changed from those proposed. In some cases, the operation of the mechanism depends on the level of agreement desired.

Thresholds may often be negotiated simply between one user and the provider, where that user wishes to be informed when QoS reaches a certain level but the provider's capability also has to be taken into account.

a) Single-parameter negotiation – Connection-wide

1) The initiating user supplies a proposed value P.

2) The provider may refuse the request. If the provider does not refuse the request, for each responding user it may select a new proposed value P_i' which is not better than the initiator-proposed value. (These new values may differ between responding users, since the capacity of the provider may vary from responding user to responding user.) Thus, for all responding users R_i, $P_i' \le P$. The provider supplies the proposed values to the responding users.

3) Each responding user may refuse the request, in which case it takes no further part in the negotiation. If a responding user does not refuse the request, it may select a new proposed value V_i which is not better than the value proposed by the provider. Thus for all responding users R_i, $V_i \le P_i' \le P$.

4) The provider shall select the lowest of the values returned by the responding users, $V = \min V_i$.

5) The selected value V is returned to the initiating user and to all the responding users. It becomes the "agreed" value, and is such that $V = \min V_i \le V_i \le P_i' \le P$.

The mechanism is illustrated in Figure 7-2.

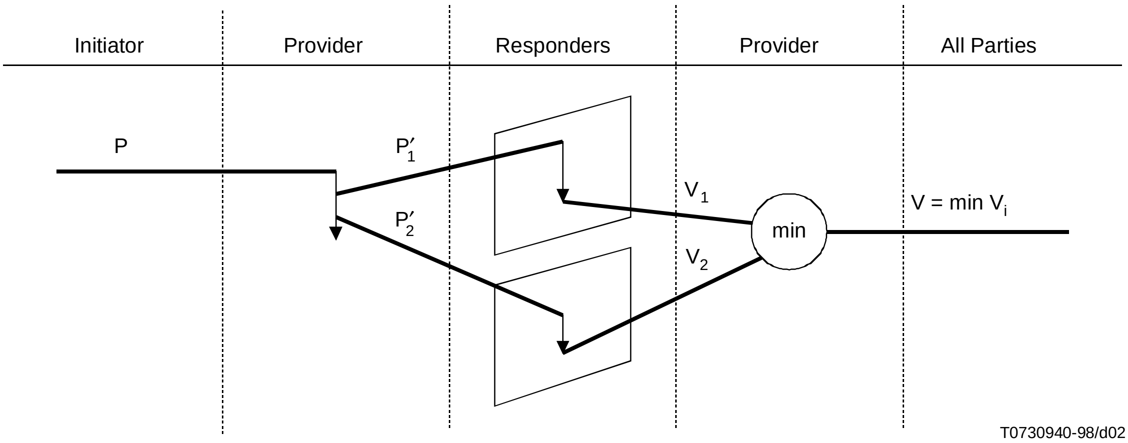


Figure 7-2 – Single-parameter negotiation (connection-wide)

b) Bounded negotiation of a low limit or low threshold – Connection-wide

1) The initiating user specifies a desired operating range by supplying a lower bound L and an upper bound U, where $L \le U$. L is its proposed low limit or low threshold value.

2) The provider may refuse the request if it knows it cannot be met, i.e. if it cannot support at least the lower bound value L. If the provider does not refuse the request, but cannot operate over the full range proposed by the initiating user, it may determine a new reduced value U_i' for the upper bound for each responding user R_i individually: this reduced value may not be worse than the lower bound. Thus $L \le U_i' \le U$ for all i. (It is also possible that the provider may choose to operate internally at a higher quality, but it does not signal this fact to the responding user.)

   NOTE 3 – It may be appropriate for the provider to propose different upper bounds to different responding users because of different provider capabilities in different regions. The provider is not required to perform an initial arbitration to determine one upper bound common to all responding users, because at this stage it is not known which responding users will wish to participate in the connection, nor the values they would wish to propose in response.

   The provider may not alter the lower bound L. The new upper bound U_i' and the lower bound L are supplied to each responding user R_i.

3) Each responding user may refuse the request, in which case it takes no further part in the negotiation. If it accepts, it may increase the lower bound to a new value L_i' within the range up to the upper bound U_i' supplied by the provider.

   Thus for each responding user R_i, $L \le L_i' \le U_i' \le U$.

   The new lower and upper bound values L_i' and U_i' are returned to the provider.

4) The provider examines the values returned from each responding user. Its behaviour will depend upon the level of agreement that is being negotiated.

Compulsory or guaranteed level of agreement

The provider must select a final connection-wide QoS value not worse than the highest lower bound of the responding users (L'_{max} = max L_i'), yet it must be capable of operating at that value to all responding users. The possibility exists that that highest lower bound L'_{max} will be greater than its operating capability, as expressed by the upper bound U_i', to one or more responding users; in such a case, some responding users must be excluded so as to leave a feasible operating region between the highest lower bound of the remaining responding users and the lowest of its remaining upper bounds.

Thus, it is a requirement for a feasible region that $L'_{max} \le U'_{min}$ and responding users may need to be removed from the connection until this constraint is satisfied.

Then the provider selects the connection-wide value V within the range, i.e. such that $L'_{max} \le V \le U'_{min}$. Typically V will be close to L'_{max}.

Best-efforts level of agreement

The provider attempts to satisfy the same constraints as in the cases of compulsory or guaranteed levels of agreement, but does not exclude responding users if the constraints cannot all be satisfied. If there is a feasible region, i.e. $L'_{max} \le U'_{min}$, the connection-wide value V selected by the provider will satisfy $L'_{max} \le V \le U'_{min}$ and typically be close to L'_{max}.

5) The selected value V is returned to the initiating user and to all (remaining) responding users. It becomes the "agreed" value. Except in the case of best-efforts level of agreement, this meets the requirements of all (remaining) parties since for all remaining responding users R_i:

$$L \le L_i' \le L'_{max} \le V \le U'_{min} \le U_i' \le U$$

The mechanism is illustrated in Figure 7-3.

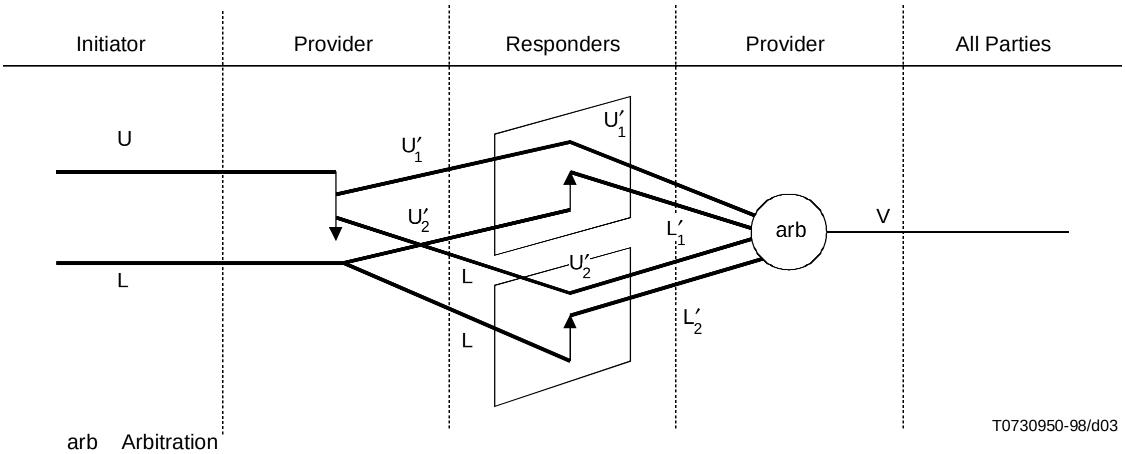


Figure 7-3 – Bounded negotiation of a low value (connection-wide)

c) Bounded negotiation of a high limit or high threshold – Connection-wide

1) The initiating user specifies a desired operating range by supplying a lower bound L and an upper bound U, where $L \le U$. U is its proposed high limit or high threshold value.

2) The provider may refuse the request if it knows it cannot be met, i.e. if it cannot support at least the lower bound value L. If the provider does not refuse the request, but cannot operate over the full range proposed by the initiating user, it may determine a new reduced value U_i' for the upper bound for each responding user R_i individually: this reduced value may not be worse than the lower bound. Thus $L \le U_i' \le U$ for all i. (It is also possible that the provider may choose to operate internally at a higher quality, but it does not signal this fact to the responding user.)

NOTE 4 – It may be appropriate for the provider to propose different upper bounds to different responding users because of different provider capabilities in different regions. The provider is not required to perform an initial arbitration to determine one upper bound common to all responding users, because at this stage it is not known which responding users will wish to participate in the connection, nor the values they would wish to propose in response.

The provider may not alter the lower bound L. The new upper bound U_i' and the lower bound L are supplied to each responding user R_i.

3) Each responding user may refuse the request, in which case it takes no further part in the negotiation. If it accepts, it may decrease the upper bound to a new value U_i'', within the bounds L and U_i' supplied by the provider.

   Thus for each responding user R_i, $L \le U_i'' \le U_i' \le U$.

   The lower and new upper bound values L and U_i'' are returned to the provider.

4) The provider selects the final connection-wide QoS value $V = \min U_i''$.

5) The selected value V is returned to the initiating user and to all responding users. It becomes the "agreed" value. This meets the requirements of all parties since for all responding users R_i:

$$L \le V = U''_{min} \le U_i'' \le U_i' \le U$$

The mechanism is illustrated in Figure 7-4.

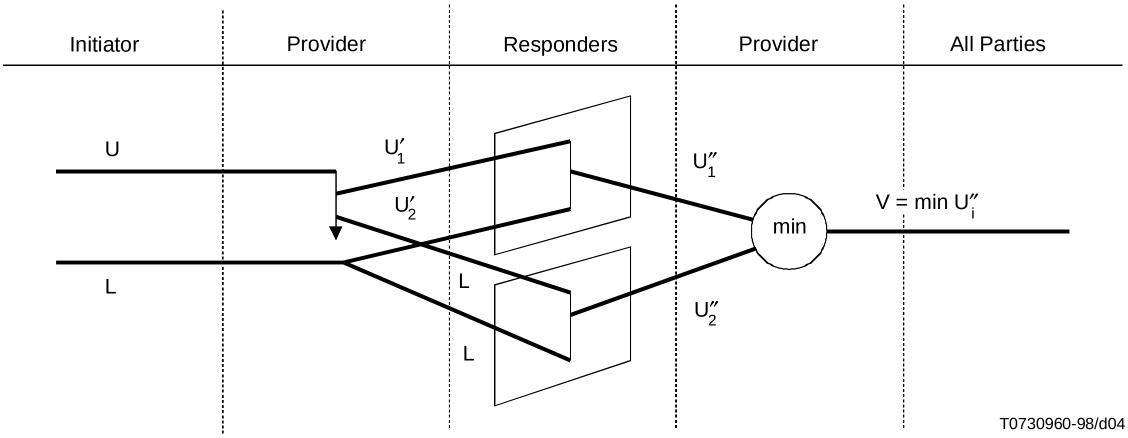


Figure 7-4 – Bounded negotiation of a high value (connection-wide)

d) Bounded negotiation of an operating target – Connection-wide

1) The initiating user specifies a desired operating range by supplying a lower bound L and an upper bound U, where $L \le U$.

2) The provider may refuse the request if it knows it cannot be met, i.e. if it cannot support at least the lower bound value L. If the provider does not refuse the request, but cannot operate over the full range proposed by the initiating user, it may determine a new reduced value U_i' for the upper bound for each responding user R_i individually: this reduced value may not be worse than the lower bound. Thus $L \le U_i' \le U$ for all i. (It is also possible that the provider may choose to operate internally at a higher quality, but it does not signal this fact to the responding user.)

   NOTE 5 – It may be appropriate for the provider to propose different upper bounds to different responding users because of different provider capabilities in different regions. The provider is not required to perform an initial arbitration to determine one upper bound common to all responding users, because at this stage it is not known which responding users will wish to participate in the connection, nor the values they would wish to propose in response.

   The provider may not alter the lower bound L. The new upper bound U_i' and the lower bound L are supplied to each responding user R_i.

3) Each responding user may refuse the request, in which case it takes no further part in the negotiation. If it accepts, it may increase the lower bound to a new value L_i' and it may decrease the upper bound to a new value U_i'', within the bounds L and U_i' supplied by the provider.

   Thus, for each responding user R_i, $L \le L_i' \le U_i'' \le U_i' \le U$.

   The new lower and new upper bound values L_i' and U_i'' are returned to the provider.

4) The provider examines the values returned from each responding user. The level of agreement that is being negotiated is best-efforts (since the others do not apply to operating targets).

The provider selects a final connection-wide QoS value V. If there is a feasible operating region within the ranges returned by all responding users, i.e. if the highest lower bound L'_{max} is less than or equal to the lowest upper bound $U''_{min} = \min U_i''$, then V is selected in the feasible region, so that $L'_{max} \leq V \leq U''_{min}$. However, this may not be possible.

5) The selected value V is returned to the initiating user and to all responding users. It becomes the "agreed" value.

The mechanism is illustrated in Figure 7-5.

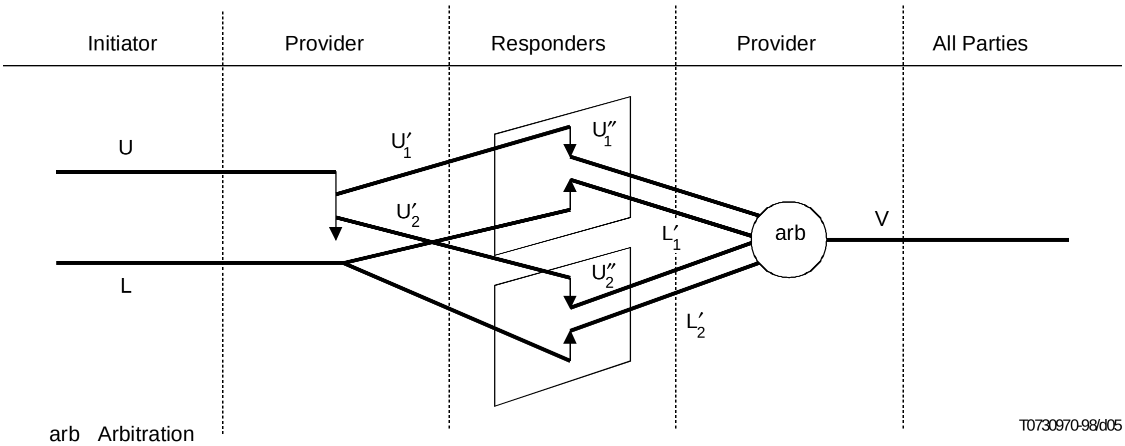


Figure 7-5 – Bounded negotiation of an operating target (connection-wide)

e) Combined negotiation of lower and upper limits – Connection-wide

This mechanism differs from the preceding ones in that it is used to negotiate two values – a lower limit and an upper limit – whereas the others are used to negotiate single values.

1) The initiating user proposes a lower limit L and an upper limit U, where $L \leq U$.

2) The provider may refuse the request if it knows it cannot be met, i.e. if it cannot support at least the lower bound value L. If the provider does not refuse the request, but cannot operate over the full range proposed by the initiating user, it may determine a new reduced value U_i' for the upper bound for each responding user R_i individually: this reduced value may not be worse than the lower bound. Thus $L \leq U_i' \leq U$ for all i. (It is also possible that the provider may choose to operate internally at a higher quality, but it does not signal this fact to the responding user.)

NOTE 6 – It may be appropriate for the provider to propose different upper bounds to different responding users because of different provider capabilities in different regions. The provider is not required to perform an initial arbitration to determine one upper bound common to all responding users, because at this stage it is not known which responding users will wish to participate in the connection, nor the values they would wish to propose in response.

The provider may not alter the lower bound L. The new upper bound U_i' and the lower bound L are supplied to each responding user R_i.

3) Each responding user may refuse the request, in which case it takes no further part in the negotiation. If it accepts, it may increase the lower bound to a new value L_i' and it may decrease the upper bound to a new value U_i'', within the bounds L and U_i' supplied by the provider.

Thus for each responding user R_i, $L \leq L_i' \leq U_i'' \leq U_i' \leq U$.

The new lower and new upper bound values L_i' and U_i'' are returned to the provider.

4) The provider examines the values returned from each responding user. Its behaviour will depend upon the level of agreement that is being negotiated.

Compulsory or guaranteed level of agreement

The provider must select a final connection-wide QoS lower limit L_F and a final connection-wide QoS upper limit U_F such that L_F is not worse than the highest lower bound $L'_{max} = \max L_i'$ and U_F is not better than the lowest upper bound $U''_{min} = \min U_i''$.

Thus, it is a requirement for a feasible region that $L'_{max} \leq U''_{min}$, and responding users may need to be removed from the connection until this constraint is satisfied.

Then the provider selects the connection-wide values L_F and U_F such that $L'_{max} \leq L_F \leq U_F \leq U''_{min}$. Typically L_F will be close to L'_{max} and U_F will be close to U''_{min}.

Best-efforts level of agreement

The provider attempts to satisfy the same constraints as in the cases of compulsory or guaranteed levels of agreement, but does not exclude responding users if the constraints cannot all be satisfied. If there is a feasible region, i.e. if $L'_{max} \leq U''_{min}$, the connection-wide values L_F and U_F selected by the provider will satisfy $L'_{max} \leq L_F \leq U_F \leq U''_{min}$ and typically L_F will be close to L'_{max} and U_F will be close to U''_{min}.

5) The selected values L_F and U_F are returned to the initiating user and to all (remaining) responding users. They are the "agreed" values. Except in the case of best-efforts level of agreement, this meets the requirements of all (remaining) parties since for all remaining responding users R_i:

$$L \leq L_i' \leq L'_{max} \leq L_F \leq U_F \leq U''_{min} \leq U_i'' \leq U_i' \leq U$$

The mechanism is illustrated in Figure 7-6.

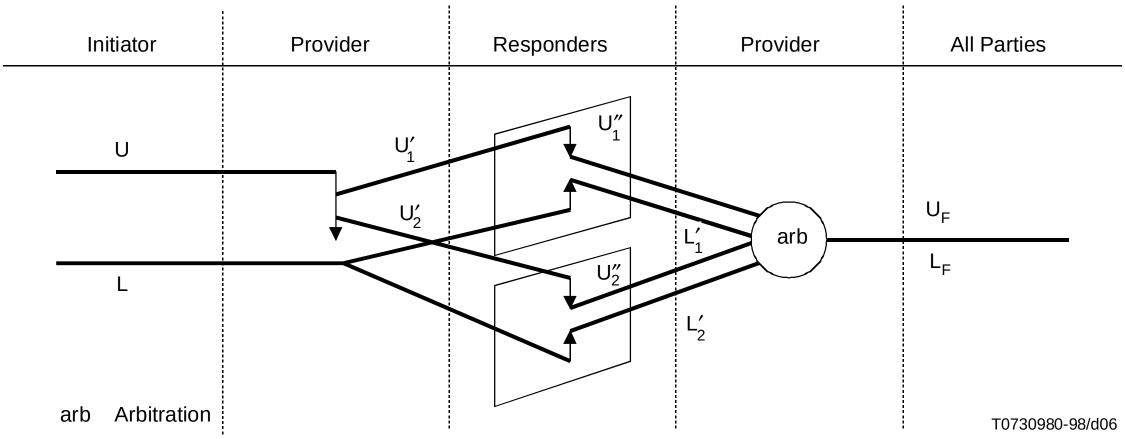


Figure 7-6 – Combined negotiation of lower and upper limits (connection-wide)

7.1.2.2 Use of QoS filters in 1 × N multicast

A complementary approach to the problem of dealing with receivers with different properties, in different environments, is to use QoS filters. A QoS filter can support different levels of QoS for different receivers in the same multipeer group. Filters are particularly valuable in dealing with continuous media information. The differences between receivers may relate to the end systems (compression boards, processing power, display devices, etc.), networks (throughput, delay, error rates, etc.) or to the user applications (quality required, frame rates, etc.).

The QoS negotiation mechanisms described in 7.1.1.1 and 7.1.2.1 can be operated in configurations where filters are also present. It is possible for a filter mechanism to act as a proxy participant in connection-wide negotiations in order to allow a receiver to join a multicast connection it would not otherwise be able to.

Subclause 8.3 discusses specific types of QoS filters that can support different values of QoS characteristics for different receivers in a multipeer group. QoS filters can be applied where different receiver-selected characteristics have been negotiated by receivers in the same 1 × N multicast connection. QoS filters enable a high level of QoS to be delivered to certain receivers, while supplying filtered information to other receivers with lower QoS requirements in the same 1 × N multicast connection. In such cases, knowledge of the properties of available filters needs to be taken into account during QoS negotiation.

The source participant may not be aware that QoS filters are being used. The source would supply a data flow at a single QoS level, while the N receivers receive either the unfiltered flow or filtered flows at different QoS levels.

7.1.3 Negotiating QoS in M × N multicast

The term M × N multicast is used to denote a multicast communication between N stations, in which M (≤ N) are intending to transmit multicast to the N. These M stations are termed "focal stations". Depending on the particular case, the N stations may be able to transmit in unicast mode to other stations. The QoS negotiated is that relating to the M multicasts.

One of the focal stations is designated the "owner" of the multicast. Typically, the owner station has a special role in initiating and terminating multicast establishment procedures, in order to ensure that the group membership rules are correctly applied and that the procedures terminate correctly. Ownership is thus a management concept.

The division between owner and non-owner stations, which concerns management roles, cuts across the division between users and provider, which reflects use and provision of communications services. In particular, an owner station will in general perform both user-level and provider-level functions, and distinctions between owner and non-owners may appear in both user-level and provider-level protocols.

NOTE – Mechanisms for changing owners are beyond the scope of this Recommendation | Technical Report.

Two kinds of procedure can be used to negotiate QoS characteristics in M × N multicast:

- one composed of superimposed 1 × N negotiations;
- a single, simultaneous M × N negotiation.

Mechanisms of these two kinds are defined in the Enhanced Communications Transport Service and Protocol, ECTS (ITU-T Rec. X.605 | ISO/IEC 13252) and ECTP (under development). ECTS and ECTP define:

- a "Step-Wise Arbitration" procedure, which is composed of superimposed 1 × N negotiations, and can make use of the mechanisms defined in 7.1.2;
- an "Owner Arbitration" procedure, in which the owner station controls a single, simultaneous M × N negotiation.

7.2 Resource allocation mechanisms

As defined in the QoS Framework, the QoS establishment phase includes mechanisms for resource allocation. One such mechanism is Resource Reservation Protocol (RSVP). A reference is given in 5.5.1.

7.3 Initialization mechanisms

As defined in the QoS Framework, the QoS establishment phase includes the initialization of operational phase mechanisms. Typically this is achieved by local means that are not subject to standardization.

8 Methods and mechanisms for the operational phase

This clause identifies sources of methods and mechanisms for the QoS operational phase, as defined in the QoS Framework.

Mechanisms for the operational phase include:

- monitoring mechanisms;
- maintenance mechanisms;
- filters;
- enquiry mechanisms;
- alert mechanisms.

8.1 Monitoring mechanisms

As noted in the QoS Framework, monitoring mechanisms can be provided by general-purpose management techniques, such as those standardized in OSI management. References are given in 5.1.2.

8.2 Maintenance mechanisms

The purpose of maintenance mechanisms is to endeavour to meet desired or agreed levels of QoS. This may be achieved by a variety of means, used singly or in combination. Some categories of maintenance mechanisms are discussed in the QoS Framework, namely:

- resource allocation;
- admission control;
- tuning.

Subclause 8.2.1 defines particular maintenance mechanisms developed to meet QoS requirements on time windows.

8.2.1 QoS management mechanisms for time-critical communications

8.2.1.1 Introduction

This subclause presents definitions of QoS management mechanisms to support requirements for complete delivery of PDUs to their destinations within time constraints. These mechanisms are used in Network Management for Time Critical Communications Systems (TCNM).

The management mechanisms use:

- an identification function, which identifies an incoming PDU as a time-critical one and extracts from the PDU the required "completion time" (T_c) before which the PDU must be delivered to its destination;
- an information base, which contains expected transfer-times for forwarding PDUs via specified routes to their destinations;
- an evaluation and processing function that can calculate an expected completion time, compare this calculated expected completion time with the required completion time attached to the PDU, determine whether the PDU can be delivered within the required completion time and, on that basis, decide whether to transfer the PDU or to abort the PDU transfer by discarding the PDU.

These mechanisms can be applied at all relevant service boundaries in a time-critical communications network system in order to satisfy time window requirements, and to provide the appropriate degree of load-shedding.

8.2.1.2 Management mechanisms

Three independent mechanisms are defined in this subclause. Various combinations of these three mechanisms may be used in practice, but such combined types are not defined here. The mechanisms are closely related to the Data Time Validity QoS characteristic. Specifically, in Time Critical Communications Systems, any data that could not be delivered within the specified time window is considered to be no longer of value to applications processes, and should be discarded in order to increase the efficient use of communications facilities.

M.1 PDU discard and notification mechanism

1) When the identification function receives a PDU, it examines part of the content of the PDU in order to determine whether the PDU is a time-critical one. If the PDU is a time-critical PDU, the identification function extracts the required time for completion of delivery to its destination (T_c) and the specified transfer route contained within the PDU.
2) The evaluation and processing function obtains from the information base the expected transfer-time necessary to forward a PDU via the specified route to its destination. The function then calculates the expected completion time over the specified route by adding the expected transfer-time to the current time, and compares the value obtained with the required completion time T_c. If delivery cannot be guaranteed, then the PDU is discarded.
3) If it decides to discard the PDU, the function generates a notification PDU reporting the discard condition and sends it to the initiator or original sender of the discarded PDU.

M.2 PDU discard and notification mechanism with dynamic priority change

This mechanism differs from the M.1 PDU discard and notification mechanism defined above in that it allows a dynamic change in the priority level attached to a time-critical PDU.

1) When the identification function receives a PDU, it examines part of the content of the PDU in order to determine whether the PDU is a time-critical one. If the PDU is a time-critical PDU, the identification function extracts the required time for completion of delivery to its destination (T_c), the specified transfer route contained within the PDU and the priority level of protocol processing associated with the content of this PDU.

2) The evaluation and processing function obtains from the information base the expected transfer-time necessary to forward a PDU via appropriate route(s) to its destination. (The route to be taken may be indicated in the PDU, or alternatively the choice may be left open.) The function then calculates the expected completion time(s) over the appropriate route(s) by adding expected transfer-time(s) to current time, and compares the value(s) with the required completion time T_c. If delivery cannot be guaranteed with the given priority level, then the evaluation and processing function determines whether delivery within the required time can be achieved by increasing the priority level to faster protocol processing. If it can, the priority level is increased as necessary. If delivery cannot be guaranteed, then the PDU is discarded.

3) If it decides to discard the PDU, the function generates a notification PDU reporting the discard condition and sends it to the initiator or original sender of the discarded PDU.

M.3 PDU discard and notification mechanism with dynamic route change

This mechanism differs from M.1 PDU discard and notification and M.2 PDU discard and notification with dynamic priority change in that it allows dynamic change in transfer route for a time-critical PDU.

1) When the identification function receives a PDU, it examines part of the content of the PDU in order to determine whether the PDU is a time-critical one. If the PDU is a time-critical PDU, the identification function extracts the required time for completion of delivery to its destination (T_c) and the specified transfer route contained within the PDU.

2) The evaluation and processing function obtains from the information base the expected transfer-time necessary to forward a PDU via appropriate route(s) to its destination. (The route to be taken may be indicated in the PDU, or alternatively the choice may be left open.) The function then calculates the expected completion time(s) over the appropriate route(s) by adding expected transfer-time(s) to current time, and compares the value(s) with the required completion time T_c. If delivery cannot be guaranteed with the given route, then the evaluation and processing function determines whether delivery within the required time can be achieved by choosing a faster route. If it can, the route is changed. If delivery cannot be guaranteed, then the PDU is discarded.

3) If it decides to discard the PDU, the function generates a notification PDU reporting the discard condition and sends it to the initiator or original sender of the discarded PDU.

8.3 Filters

A QoS filter is defined as a mechanism that transforms data in order to alter some properties related to QoS, for example the QoS needed to transfer the data or the value of one or more QoS characteristics of the service provider.

QoS filters can be used in peer-to-peer communications and multipeer communications. QoS filters are suited to data containing information that can be discarded, is resilient to loss or is such that a number of levels of QoS can be made available from the same original data.

This clause primarily describes the use and effects of filters during the transmission of media; negotiation of these mechanisms is discussed in 7.1.2.2.

8.3.1 Generic filter types

QoS filters are grouped into three types, as follows.

8.3.1.1 Intelligent media-discarding filter

Intelligent media-discarding is a means of altering the QoS properties of a data flow by dynamically removing data from the ongoing flow. The data must be discarded intelligently to ensure the resultant data is not corrupted. Examples include a filter which removes colour information from a video flow leaving just grey-scale video.

8.3.1.2 Translation filter

Translation filtering is a means of altering the QoS properties of a data flow by converting the data in the ongoing flow in some way. For example, data may be compressed or decompressed by such a filter.

8.3.1.3 Implicit filter

Implicit filtering is a means whereby separate components of a data stream are transmitted in data flows on separate connections. The QoS properties of the overall data stream can be altered by establishing and breaking the various connections that comprise the whole data stream. An example is the use of the scaleable syntax.

8.3.2 Objectives of filtering

QoS filters are used to alter one or more QoS characteristics of a data flow. A filter may be used to make required changes to the following QoS characteristics.

8.3.2.1 Throughput

A QoS filter can have significant effects on data throughput. By discarding PDUs destined for certain participants, it may be possible to satisfy throughput constraints imposed by some participants or by the provider of the communications service in certain regions. Media-discarding filters and filters performing compression are primarily intended to reduce the throughput requirements that particular participants may have for particular data flows.

8.3.2.2 Loss sensitivity

A QoS filter can be used to translate data with high loss sensitivity to a form with improved error resilience (low loss sensitivity). This may be performed by implementing error detection and recovery schemes.

8.3.2.3 Delay

The end-to-end delay may be decreased by reducing data throughput requirements and releasing network resources, as a result of a reduction of incurred queuing in the network fabric.

8.3.2.4 Jitter

Jitter reduction may be achieved by resynchronizing time-stamped PDUs as part of the operation of a filter. A QoS filter that translates a variable bit-rate data flow to a constant bit-rate data flow is another type of translation filter that can be used to reduce jitter.

8.3.3 Side effects of filtering on QoS characteristics

The side effects of applying QoS filters may be that some QoS characteristics are adversely effected. Such effects include the following.

8.3.3.1 Throughput

QoS filters that add information to a data flow increase the requirement for raw throughput of data, which if not accommodated may lead to a decrease in the throughput perceived by users. Such filters include decompression filters and those that increase error resilience by adding error detection and correction data.

8.3.3.2 Loss sensitivity

Dependent on the type of media and encoding method, discarding information and removing redundancy may increase the loss sensitivity of a data flow.

8.3.3.3 Delay

Performing intensive QoS filter operations, as in translation filtering, will increase the experienced end-to-end delay. The effect on delay is dependent on the filter implementation and filter processing device used. As noted above, the end-to-end delay may in some cases be decreased as a result of a reduction of throughput requirement.

8.3.3.4 Jitter

The jitter of a data flow will be affected to a degree dependent on the type of media and encapsulation strategy used. Performing filtering on larger PDUs will cause larger delays, while filtering operations on smaller PDUs will cause smaller delays.

8.4 Enquiry mechanisms

As noted in the QoS Framework, enquiry mechanisms can be provided by general-purpose management techniques, such as those standardized in OSI or SNMP management. References are given in 5.1.2.

8.5 Alert mechanisms

As noted in the QoS Framework, alert mechanisms can be provided by general-purpose management techniques, such as those standardized in OSI or SNMP management. References are given in 5.1.2.

9 QoS verification methods

9.1 Introduction

The QoS Framework examines QoS verification at different stages of the lifecycle of the implementation of a service. This clause defines some QoS verification methods for these stages.

9.2 QoS verification at the design stage

The service design phase of the service lifecycle starts from a set of requirements on the service and derives a complete system design. The design process is characterized by stepwise refinement, and involves a formalization of informal service requirements. At the end of the design phase, there is value in verification that the system (resulting from the design) actually meets the stated service requirements.

Service requirements relate to the expected behaviour of the service, but also comprise non-functional aspects. Besides the sequential order of events, which determines the interaction of the service users with the service, there are also requirements on the timeliness, capacities or reliability for certain events to happen at a certain state.

Figure 9-1 shows the main steps of a method for QoS verification during the design phase for a QoS enhanced design. The service designer starts from a set of requirements comprising functional and non-functional aspects. In the first step, the requirements are formalized to provide an abstract service model.

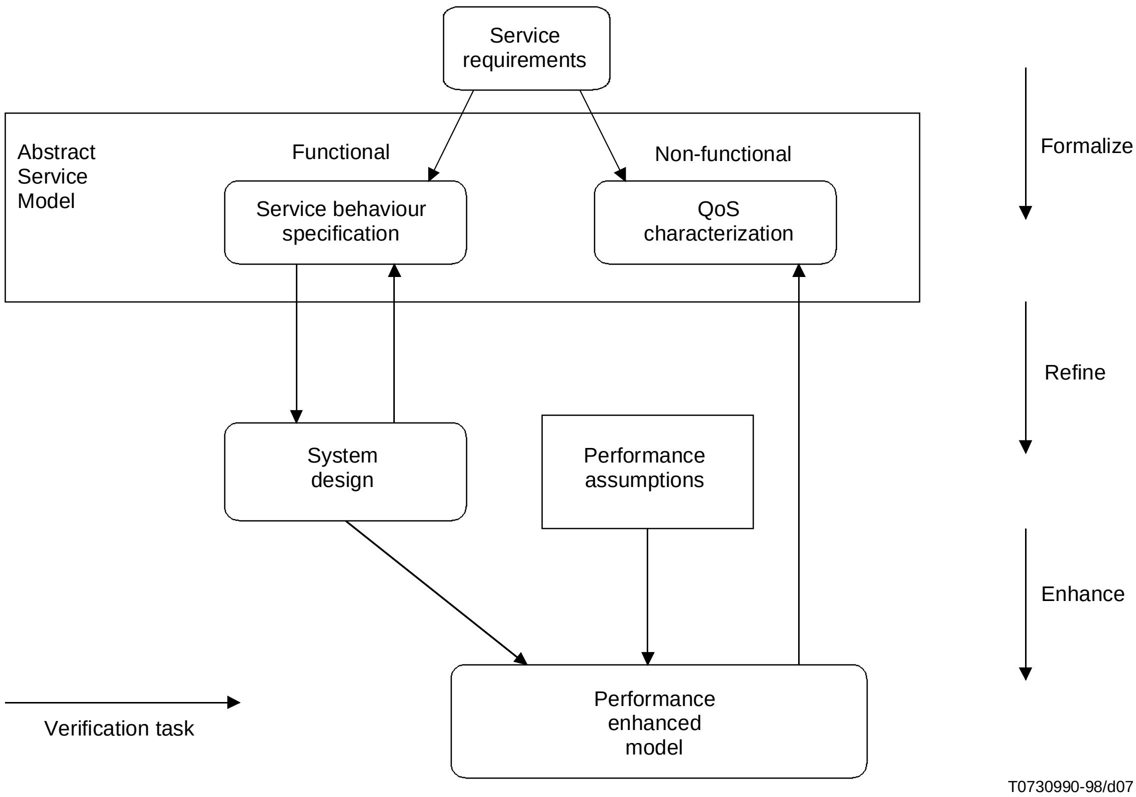


Figure 9-1 – QoS-based service design process

In this method, a hybrid approach is used that separates the behavioural aspects of the service from the non-functional requirements relating to QoS. This separation of concerns is motivated by the different level of abstraction on which these types of requirements are expressed [e.g. time requirements generally relate to entire (long) sequences of events instead of being an attribute of a single event in relation to its single predecessor only]. By having separate specifications, it is also easier to do modifications that may arise in iterations during the design cycle.

The abstract service behaviour is specified using a Formal Description Technique (FDT). By adding details (e.g. functional decomposition), the specification is then refined in correctness by preserving transformation steps that lead to a complete design. The consistency between the abstract service behaviour and its design solution can be verified by proving the equivalence of the observational behaviour between the specifications.

Along with the formalization of the functional behaviour, there should be a formalization of the QoS-related requirements. In this process, the relevant QoS characteristics of the service have to be identified and their values must be selected such that they match the users' expectations of the service quality. The functional decomposition in the refinement process is paralleled by a decomposition of the global QoS requirements into QoS requirements of service components.

With a detailed functional system specification, the designer is able to add performance information to the model. The performance assumptions are based on experience and on measurements of the underlying systems. Those values, of course, depend strongly on the chosen technology for the implementation. For some performance values there may be only estimates available that need refinement in further interaction cycles until the observed system performance is matched. The result of the process is a performance enhanced design model, from which parameter values on the supported QoS can be derived which allow a comparison with the required QoS constraints.

For different types of specifications used within the design process, appropriate notations are needed. For functional behaviour, the standardized FDTs may be suitable. However, it is not a requirement that the FDT should be standardized.

QoS description requires a notation that allows for defining the QoS characteristics. The notation is used to specify the constraints on these characteristics and to describe procedures for their measurement. The QoS characterization comprises the definition of boundary values, statistical aspects and performance conditions on the environment, under which certain QoS parameter values are required to hold, etc.

9.3 QoS verification at the testing stage

The objective of QoS verification at the testing stage is to verify that a service implementation fulfils its QoS requirements. The approach taken here is to use the methods of Protocol Conformance Testing (PCT) [see ITU-T Rec. X.290 series | ISO/IEC 9646 (all parts)] for QoS testing.

Although the ITU-T Rec. X.290 series | ISO/IEC 9646 (all parts) has been extended to cover multiparty testing, it is possible that some generalization or enhancement of approach will be needed. Instead of a layered protocol stack, the tester has to interact with a multiparty service offered on a distributed platform which makes use of various resource services (e.g. communication services, data bases). Therefore, a distributed test environment is needed that is capable to measure, control and evaluate QoS.

The testing of behaviour is mainly based on a stimulus-response principle, where a sequence of controlling and observed events have to be checked for conformity. QoS testing requires, in addition, measurement procedures for QoS characteristics. During the test operation it should be checked that the tested characteristics are within the required boundary values.

For QoS measurement, clearly-defined measures of characteristics are needed and points must be defined where the measurement can be performed. Finally, a procedure must be defined how the measurement is effected.

Certain theoretically derived characteristics which rely on idealistic assumptions have to be adopted to practical approximations, e.g. availability aspects have to be measured by suitable sample sizes. Also, it is necessary that the measurement itself does not interfere to such a degree that it falsifies the measured characteristics.

For QoS characteristics, we can distinguish between directly measurable characteristics (e.g. time delay) and derived characteristics. A typical example for the latter is packet loss, which is computed from the difference between sent and received data units. The measurement technique for a certain characteristic may require support from the system and its environment. A delay can be defined as round-trip delay, which requires only a single point of measurement and local time, but depends on the cooperating responding behaviour at the receivers' end, while simple delays require synchronized clocks between sender and receiver and either a measurement point at the sender and the receiver side or a transmission of a time stamp on the sent data with only one measurement point at the receiver.

Figure 9-2 presents a distributed architecture for QoS Testing. The Service Under Test (SUT) must have access to its required resource services. The role of the service users is played by a set of local testers. There must be points of measurement between SUT and the local testers for the measurement of the QoS provided by the service, and there must be measurement points between SUT and resource services for the measurement of the QoS received from the resource services. This information might be obtained from Management Information Bases (MIB), too. For the overall coordination of the test scenario, there is a central component controlling the distributed elements of the test environment. It is necessary to have a component responsible for the measurement of QoS characteristics of the SUT, and for the control of the QoS of the resource services (environment constraints). Typical elements that form part of the Control and Measurement Unit include traffic and error generators.

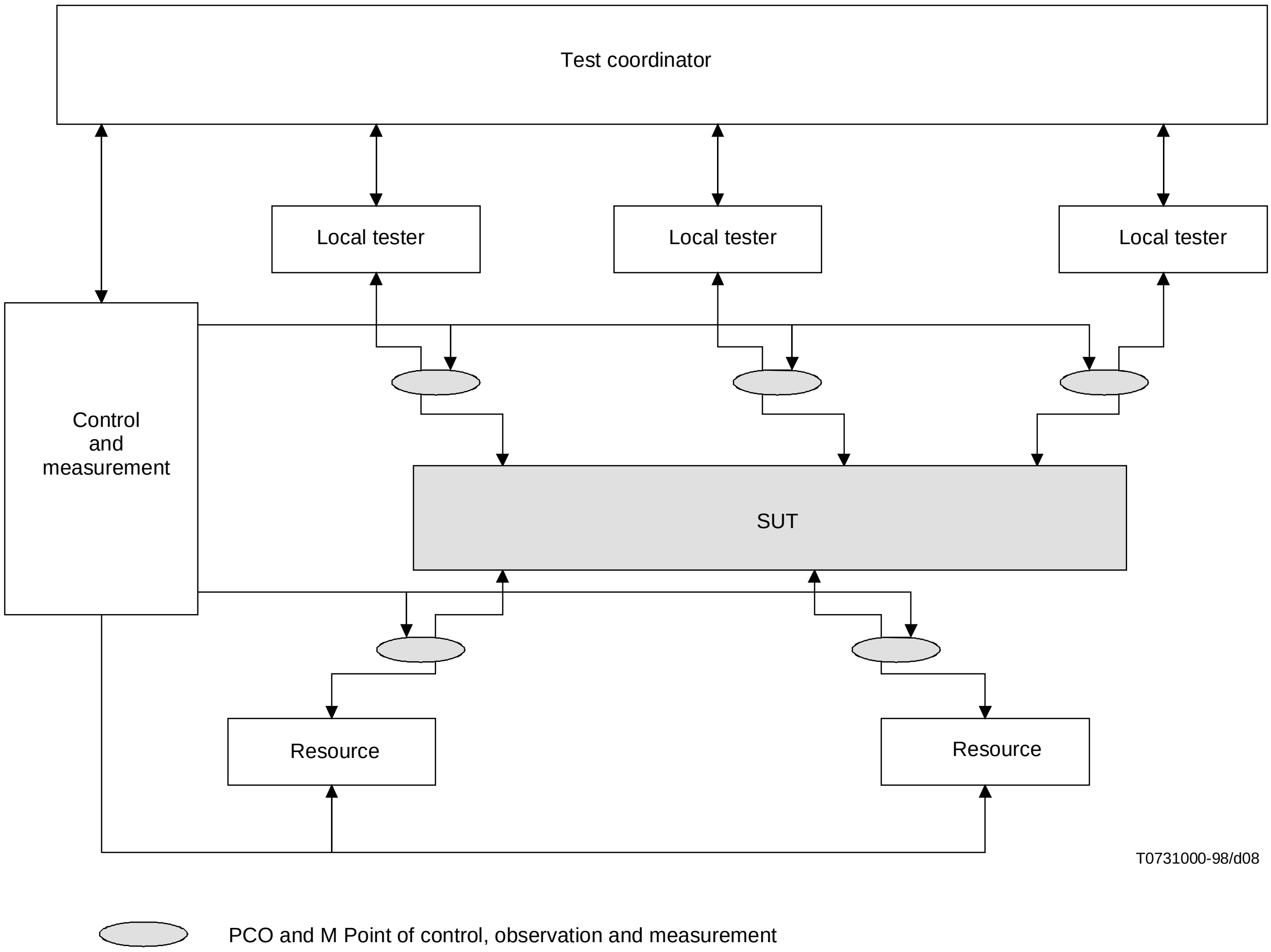


Figure 9-2 – QoS-based service testing architecture

9.4 QoS verification at the operation stage

If a service offers QoS only on a best effort basis, it may be not capable of guaranteeing a certain level of QoS and, consequently, the resulting QoS as presented to the user may differ from the required value, e.g. the intermediate services that make use of basic resource services are engineered to optimize the utilization of the scarce resources and may require guaranteed QoS; e.g. constant bit rate compression service is favoured to make optimal use of a fixed allocated bandwidth.

For a service offering a guaranteed level of QoS, there should be continuous monitoring of the QoS provided to the user and the QoS received from the underlying resource services. When certain threshold values are reached, actions according to a predefined policy (e.g. reservation of additional resources) could be taken and internal management mechanisms used to keep to the contracted level.

10 References to this Recommendation | Technical Report

Recommendations, International Standards or Technical Reports that reference methods or mechanisms defined in this Recommendation | Technical Report should state that they are "consistent" with this Recommendation | Technical Report. This means that the meaning of the definitions is not altered by the referencing Recommendation, International Standard or Technical Report.

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 communications**
Series Y	Global information infrastructure
Series Z	Programming languages