



INTERNATIONAL TELECOMMUNICATION UNION

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

J.174

(02/2002)

SERIES J: CABLE NETWORKS AND TRANSMISSION
OF TELEVISION, SOUND PROGRAMME AND OTHER
MULTIMEDIA SIGNALS

IPCablecom

IPCablecom interdomain quality of service

ITU-T Recommendation J.174

ITU-T J-SERIES RECOMMENDATIONS
CABLE NETWORKS AND TRANSMISSION OF TELEVISION, SOUND PROGRAMME AND OTHER
MULTIMEDIA SIGNALS

General Recommendations	J.1–J.9
General specifications for analogue sound-programme transmission	J.10–J.19
Performance characteristics of analogue sound-programme circuits	J.20–J.29
Equipment and lines used for analogue sound-programme circuits	J.30–J.39
Digital encoders for analogue sound-programme signals	J.40–J.49
Digital transmission of sound-programme signals	J.50–J.59
Circuits for analogue television transmission	J.60–J.69
Analogue television transmission over metallic lines and interconnection with radio-relay links	J.70–J.79
Digital transmission of television signals	J.80–J.89
Ancillary digital services for television transmission	J.90–J.99
Operational requirements and methods for television transmission	J.100–J.109
Interactive systems for digital television distribution	J.110–J.129
Transport of MPEG-2 signals on packetised networks	J.130–J.139
Measurement of the quality of service	J.140–J.149
Digital television distribution through local subscriber networks	J.150–J.159
IPCablecom	J.160–J.179
Miscellaneous	J.180–J.199
Application for Interactive Digital Television	J.200–J.209

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

ITU-T Recommendation J.174

IPCablecom interdomain quality of service

Summary

This Recommendation describes a set of end-to-end Quality-of-Service (QoS) mechanisms for IPCablecom inter- and intra-domain environments.

Source

ITU-T Recommendation J.174 was prepared by ITU-T Study Group 9 (2001-2004) and approved under the WTSA Resolution 1 procedure on 13 February 2002.

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications. The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. 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 Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 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

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. 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, 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.

© ITU 2002

All rights reserved. No part of this publication may be reproduced, by any means whatsoever, without the prior written permission of ITU.

CONTENTS

Page

1	Scope	1
2	References.....	1
2.1	Normative references.....	1
2.2	Informative references.....	1
3	Terms and definitions	2
4	Abbreviations, acronyms and conventions.....	2
4.1	Abbreviations and acronyms	2
4.2	Conventions.....	3
5	Introduction	3
5.1	Solution requirements.....	3
5.2	Requirements phasing	4
5.3	General objectives	5
6	Network model	5
7	Diffserv usage in backbone	6
7.1	Media traffic	6
7.2	Signalling traffic.....	6
7.3	PHB selection and DSCP setting.....	7
7.4	PHB support by AN.....	8
7.5	Resource allocation	8
7.6	Admission control	8
8	Admission control for a single domain.....	9
8.1	Per-flow RSVP control plane	9
8.1.1	AN behaviour	10
8.1.2	Location of Diffserv edge.....	13
8.1.3	Edge router behaviour	13
8.1.4	Other terminating devices (media gateways, anonymizers, announcement servers, conference bridges)	13
8.1.5	Core router behaviour.....	14
8.1.6	Signalling latency	14
8.1.7	Pre-emption	14
8.2	Aggregate RSVP	14
8.2.1	Provisioned aggregate reservations	15
8.2.2	Dynamic aggregate reservations.....	15
8.2.3	Hierarchical aggregation	16
8.2.4	Location of aggregation points and DiffServ edge.....	16

	Page
8.3 Bandwidth broker	16
9 Admission control over multiple domains.....	17
10 Use of MPLS	18
11 Queuing and filtering.....	19
11.1 Queuing	19
11.2 Filtering	19
Appendix I – Call flow examples	20
Appendix II – Bibliography	21

ITU-T Recommendation J.174

IPCablecom interdomain quality of service

1 Scope

This Recommendation describes a set of Quality-of-Service (QoS) mechanisms for the IPCablecom project. The objective of this Recommendation is to define an architectural model for end-to-end Quality of Service for IPCablecom inter- and intra-domain environments. The Recommendation describes mechanisms for integrating IPCablecom Dynamic Quality of Service (DQoS) signalling protocols with current IP Core Network QoS models. Networks with no QoS control are out of the scope of this Recommendation. This Recommendation assumes familiarity with the IPCablecom architecture, specifically with DQoS and call signalling.

2 References

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published.

2.1 Normative references

- [1] IETF RFC 2212 (1997), *Specification of Guaranteed Quality of Service*.
- [2] IETF RFC 2474 (1998), *Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers*.
- [3] IETF RFC 2475 (1998), *An Architecture for Differentiated Service*.
- [4] IETF RFC 2998 (2000), *A Framework for Integrated Services Operation over Diffserv Networks*.
- [5] IETF RFC 3175 (2001), *Aggregation of RSVP for IPv4 and IPv6 Reservations*.
- [6] IETF RFC 3181 (2001), *Signalled Preemption Priority Policy Element*.
- [7] ITU-T Recommendation J.163 (2001), *Dynamic quality of service for the provision of real time services over cable television networks using cable modems*.

2.2 Informative references

- [8] IETF RFC 2638 (1999), *A Two-bit Differentiated Services Architecture for the Internet*.
- [9] IETF RFC 2597 (1999), *Assured Forwarding PHB Group*.
- [10] IETF RFC 2598 (1999), *An Expedited Forwarding PHB*.
- [11] IETF RFC 2702 (1999), *Requirements for Traffic Engineering Over MPLS*.
- [12] IETF RFC 3031 (2001), *Multiprotocol Label Switching Architecture*.
- [13] IETF RFC 3084 (2001), *COPS Usage for Policy Provisioning (COPS-PR)*.
- [14] ITU-T Recommendation J.171 (2002), *IPCablecom trunking gateway control protocol (TGCP)*.
- [15] MAKAM [S.] et al.: Framework for MPLS-based Recovery, *draft-ietf-mpls-recovery-frmwrk-03.txt*, July 2001.

3 Terms and definitions

This Recommendation defines the following terms:

3.1 access node: As used in this Recommendation, an Access Node is a layer two termination device that terminates the network end of the J.112 connection. It is technology specific. In J.112 Annex A, it is called the INA, while in Annex B and Annex C it is the CMTS.

3.2 endpoint: A Terminal, Gateway or MCU.

3.3 flow [IP flow]: A unidirectional sequence of packets identified by ISO Layer 3 and Layer 4 header information. This information includes source/destination IP addresses, source/destination port numbers, protocol ID. Multiple multimedia streams may be carried in a single IP Flow.

3.4 flow [J.112 flow]: A unidirectional or bidirectional flow of data packets that is subject to MAC-layer signalling and QoS assignment compliant to ITU-T Rec. J.112. Multiple multimedia streams may be carried in a single J.112 Flow.

3.5 gateway: Devices bridging between the IP/Cablecom IP Voice Communication world and the PSTN. Examples are the Media Gateway which provides the bearer circuit interfaces to the PSTN and transcodes the media stream, and the Signalling Gateway which sends and receives circuit-switched network signalling to the edge of the IP/Cablecom network.

3.6 latency: The time, expressed in quantity of symbols, taken for a signal element to pass through a device.

3.7 proxy: A facility that indirectly provides some service or acts as a representative in delivering information there by eliminating a host from having to support the services themselves.

3.8 trunk: An analog or digital connection from a circuit switch which carries user media content and may carry voice signalling (MF, R2, etc.).

4 Abbreviations, acronyms and conventions

4.1 Abbreviations and acronyms

This Recommendation uses the following abbreviations:

AF	Assured Forwarding
AN	Access Node
ATM	Asynchronous Transfer Mode
CMS	Call Management Server
COPS	Common Open Policy Service Protocol
DCS	Distributed Call Signalling
DQoS	Dynamic Quality of Service
DSCP	Differentiated Services Codepoint
EF	Expedited Forwarding
ER	Edge Router
IETF	Internet Engineering Task Force
IntServ	Integrated Services
IP	Internet Protocol
MPLS	Multiprotocol Label Switching

MTA	Media Terminal Adapter
PHB	Per-Hop Behaviour
PHS	Payload Header Suppression
PSTN	Public Switched Telephone Network
QoS	Quality of Service
RSVP	Resource reSerVation Protocol
UDP	User Datagram Protocol
VoIP	Voice over Internet Protocol

4.2 Conventions

If this Recommendation is implemented, the key words "MUST" and "SHALL" as well as "REQUIRED" are to be interpreted as indicating a mandatory aspect of this Recommendation.

The key words indicating a certain level of significance of a particular requirements that are used throughout this Recommendation are summarized below.

"MUST"	This word or the adjective "REQUIRED" means that the item is an absolute requirement of this Recommendation.
"MUST NOT"	This phrase means that the item is an absolute prohibition of this Recommendation.
"SHOULD"	This word or the adjective "RECOMMENDED" means that there may exist valid reasons in particular circumstances to ignore this item, but the full implications should be understood and the case carefully weighed before choosing a different course.
"SHOULD NOT"	This phrase means that there may exist valid reasons in particular circumstances when the listed behaviour is acceptable or even useful, but the full implications should be understood and the case carefully weighed before implementing any behaviour described with this label.
"MAY"	This word or the adjective "OPTIONAL" means that this item is truly optional. One vendor may choose to include the item because a particular marketplace requires it or because it enhances the product, for example; another vendor may omit the same item.

5 Introduction

5.1 Solution requirements

There are three basic requirements to providing end-to-end QoS for IPCablecom sessions:

- 1) provide acceptable call set-up times, comparable to those in the PSTN;
- 2) provide acceptable voice quality by providing mechanisms to guarantee sufficiently small delay, jitter, and packet loss;
- 3) ensure high quality is maintained for the entire duration of the session (e.g. block new call attempts when their completion would compromise the quality of existing calls).

In a packet-based network, the second requirement translates to: provide mechanisms to recognize IPCablecom traffic and manage scheduling and buffer allocation in each switch and router so that delay and packet loss are bounded.

The third requirement defines the need for admission control. Depending on the QoS mechanisms chosen, the challenge is to define a satisfactory method to block or admit calls or sessions based upon resource availability in the backbone.

The following are the general criteria to evaluate solutions for end-to-end QoS for IPCablecom.

- The solution should meet the three requirements above.
- The solution is manageable and implementable.
- The solution is scalable. QoS mechanisms for voice communication services must be able to grow to accommodate a large number of concurrent IPCablecom sessions without introducing undue implementation costs or complexity.
- The solution should recover gracefully when network failures occur. For example, it is probably unavoidable that some calls are dropped when a network failure occurs, but this event should not negatively influence other calls in the network.

These requirements, in particular the requirement for scalability, lead to a backbone architecture that is based on the IETF's Differentiated Services (Diffserv) approach [2], [3]. Diffserv was specifically designed as a scalable approach to delivering QoS in large backbones. Its application in the IPCablecom environment is described in the following clauses.

5.2 Requirements phasing

This Recommendation presents several approaches for providing QoS across a managed IPCablecom backbone network. Several of the approaches are complementary and based upon the resource management needs of the network operator, these approaches mechanisms may be combined to produce the desired control and management of IPCablecom resources and sessions.

The following table illustrates the feasible combinations of approaches described in clauses 7 through 10.

Approach	Required clauses
Diffserv	Clause 7 (Diffserv)
Per-Flow RSVP	Clauses 7, 8.1
Aggregate RSVP	Clauses 7, 8.1, 8.2
BW Broker	Clauses 7, 8.3

Diffserv support is **REQUIRED** for all IPCablecom backbone networks. IPCablecom devices **MUST** at a minimum support the DiffServ requirements defined in clause 7.

Per-flow RSVP requirements as defined in clauses 8 and 8.1 are **OPTIONAL**. However, if per-flow RSVP is supported, the requirements as defined in clauses 8 and 8.1 are **REQUIRED**.

Clause 8.2 describes an approach for aggregation of RSVP. If RSVP aggregation is supported, all of 8.2 is **REQUIRED**. In addition, all of the per-flow RSVP requirements as defined in clauses 8 and 8.1 are **REQUIRED**.

MPLS optimizations, as described in clause 9, are **OPTIONAL** and may be used with any of the approaches described.

5.3 General objectives

The general objectives of this Recommendation effort are to:

- define signalling mechanisms for establishment of QoS resources between ANs that are separated by a managed IP backbone network;
- define signalling mechanisms for establishment of QoS resources between ANs and other IP-Cablecom elements in the media path such as Edge Routers, Border Routers, Media Gateways, and Media Servers;
- support end-to-end Dynamic QoS sessions across managed IP backbone networks;
- define the interfaces for control and delivery of QoS between IP-Cablecom domains;
- support Network-based Call Signalling (NCS) and Distributed Call Signalling (DCS) models;
- support both layer-2 QoS signalling (J.112) and layer-3 QoS signalling (RSVP) on the access network;
- support multiple backbones with standard QoS implementations for managing scheduling and buffer allocation in switches and routers (e.g. MPLS, DiffServ, ATM, RSVP, etc.).

6 Network model

The overall IP-Cablecom network architecture is depicted in Figure 1. An IP-Cablecom backbone network consists of a general topology managed IP network that may comprise multiple administrative domains.

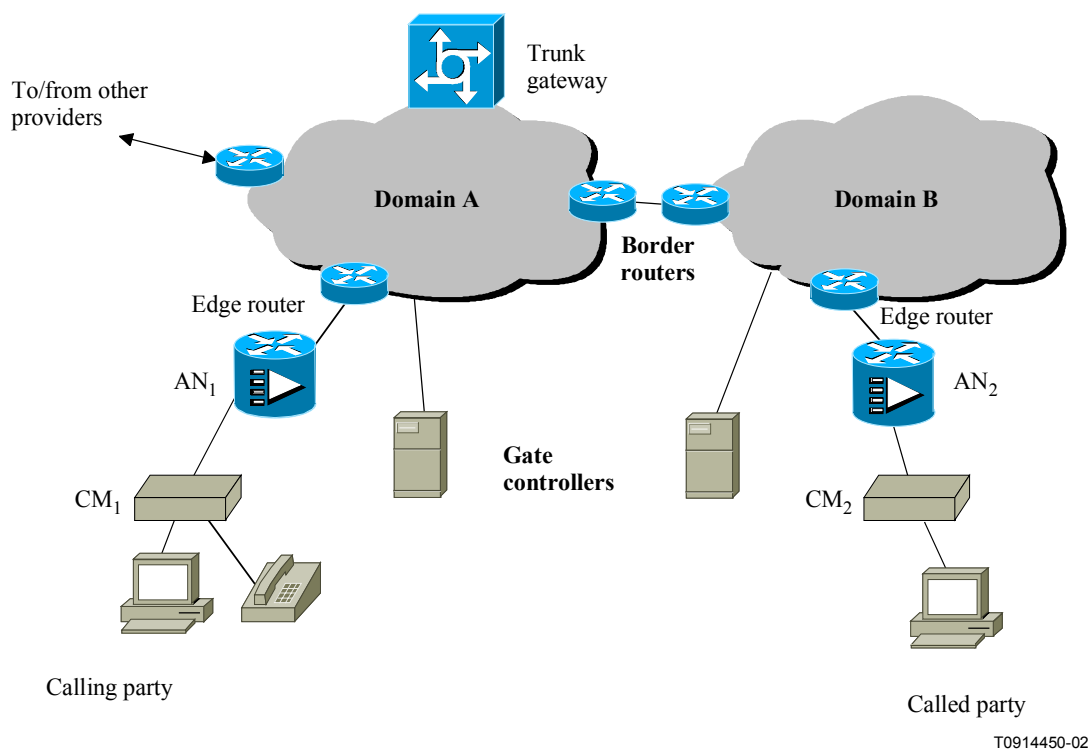


Figure 1/J.174 – Interdomain QoS architecture

The architecture assumes the existence of a service agreement between service providers that defines the level of trust between IP-Cablecom domains as well as requirements for QoS, call signalling, transport and interconnection requirements, and other such details.

The architecture also supports the transport of media and signalling between domains and may pass through one or more intermediate or transit IP networks. IPCablecom assumes that operators will have transport agreements with all transit networks to which they are directly connected.

In this architecture, we assume that DQoS signalling is used in the access network. The access portion of the network is defined to be between the Media Terminal Adapter (MTA) and the Access Node (AN), and includes the J.112/HFC network.

At a minimum, the backbone and transit IP networks are expected to be compliant with the Diffserv architecture. The backbone portion of the network is defined to be all of the IP network elements between the two ANs. This includes all edge, border, and core routers. For sessions that terminate on the PSTN, the backbone network may be further defined to include all resources between the AN and Media Gateway.

The following clauses describe a number of approaches that can be taken which offer different degrees of assurance and levels of complexity.

Note that the AN may be at the edge of the Diffserv backbone or not, depending on various factors described below. Also note that there may be intermediate devices (not shown in Figure 1) between the AN and the edge of the Diffserv backbone.

Border routers are those that sit at the boundaries between providers. They have specific roles in a Diffserv environment (such as aggregate policing and re-marking) that are discussed in more detail in the clauses that follow.

7 Diffserv usage in backbone

In this clause, we assume a simple Diffserv backbone [2], [3] with no signalling of resource requirements beyond those specified in DQoS. In this case the AN functions as the Diffserv edge device. In later clauses, we build on the foundation of a Diffserv backbone by adding signalling capabilities to control access to resources in the backbone. We assume here a common backbone for data and voice; other possibilities include using a physically or logically separate network for voice.

7.1 Media traffic

IPCablecom media traffic is defined as packets originating or terminating on a IPCablecom endpoint for which QoS has been requested using DQoS. (Note that this explicitly excludes call and QoS signalling packets such as DQoS gate coordination messages, DCS/SIP Invite messages, etc., which are discussed in 7.2). In the backbone, at least one Per-Hop Behaviour (PHB) SHOULD be dedicated for IPCablecom media traffic. This PHB MAY be EF [10], one of the AF [9] PHBs, one of the Class Selector (CS) PHBs, or a "private" PHB. The only restrictions are:

- It MUST NOT be the default (best effort) PHB.
- The only packets which are assigned this PHB SHOULD be those for which QoS was requested using DQoS.
- If an AF PHB is used, it SHOULD be Afx1, i.e. it should offer the lowest drop probability.

It is not required that all domains use the same PHB for IPCablecom media packets. It is also possible to use more than one PHB within a single domain for IPCablecom media packets, in which case it is necessary to provide some policy at the AN to determine which of the possible PHBs to use for a given packet. We return to this issue below.

7.2 Signalling traffic

Signalling traffic is defined to include call-signalling messages between IPCablecom call control elements (e.g. DCS or NCS messages, DQoS Gate Coordination messages, RSVP messages, etc.). In order to control the latency and loss experienced by signalling traffic, one or more PHBs MAY

be dedicated to IPCablecom signalling traffic. For example, the PHB CS6 has traditionally been used for routing traffic. If a PHB is dedicated to signalling messages, the following guidelines apply:

- The PHB for signalling messages **SHOULD** be distinct from the default best-effort PHB.
- The only packets which are assigned this PHB **SHOULD** be signalling messages.
- The PHB for signalling messages **MUST** be distinct from the PHB that is used for routing messages.
- The PHB for signalling messages **SHOULD** be distinct from the PHB that is used for media messages.
- The amount of traffic generated with the PHB **SHOULD** be limited.

This last guideline may be difficult to achieve. A possible approach is to limit the amount of traffic that a single user may generate that is marked with the DSCP for this PHB to a configured value. This value needs to be just large enough to accommodate the expected signalling load from the user. This limit may be enforced by policing packets bearing the designated DSCP at the AN, with excess packets being remarked "best effort". Excess packets **SHOULD NOT** be dropped. By re-marking excess packets, users are prevented from sending significant amounts of data traffic with the DSCP that is reserved for signalling. At the same time, if the offered load of signalling traffic temporarily exceeds the expected level, the excess is still transmitted into the network with a reasonable chance of timely delivery, thus avoiding serious degradation in signalling performance.

Note that the use of different PHBs for media and signalling traffic does not need to imply a relative prioritization of media over signalling, or vice versa. The intent is simply to allow resources to be allocated for media and signalling independently to meet the desired loss and latency goals for each traffic type.

7.3 PHB selection and DSCP setting

The AN is required to set or police the DSCP for IPCablecom media and signalling packets. The IPCablecom DQoS Recommendation [7] provides a means by which a Gate Controller can tell an AN which DSCP to use on a call-by-call basis, via the GATE-SET message. The AN **MUST** ensure that the DSCP for all media packets for a given call is set to the value contained in the GATE-SPEC for that call. The AN **MUST** ensure that the amount of media traffic generated for a given call that is marked with the desired DSCP does not exceed the token bucket recommendation that was provided by DQoS signalling.

Each domain **MAY** employ its own DSCP(s) for whichever PHB(s) it uses independently of other domains, as long as the choice is consistent across a single domain. If a standard PHB is used, the IETF recommended code point **SHOULD** be used as defined in [9] and [10]. If different DSCPs are used for IPCablecom media and signalling packets in neighboring domains, DSCP remarking **MUST** be performed by a border router on packets that leave one domain and enter another. The required capabilities of border routers are described below.

Routers at the domain borders **MUST** also be able to set or police the DSCP for packets which are destined to a IPCablecom endpoint and for which QoS has been requested using DQoS. In the absence of explicit signalling at domain boundaries, it is not possible to authoritatively identify IPCablecom media packets arriving at a border router on a per-flow basis. Thus, border routers must rely on the DSCP to identify such packets. For this reason, the border router **SHOULD** provide the following capabilities:

- It **SHOULD** be possible to configure the border router to impose a limit on the total amount of traffic entering the domain that is marked with a certain DSCP.
- It **SHOULD** be possible to configure the border router to modify the DSCP of traffic entering the domain. This capability is used if it is known that a different DSCP is in use

for IPCom media packets in a domain from which packets are received. That is, a router may be configured to recognize packets arriving on one interface with DSCP = x as IPCom media packets and then transmit them on another interface with DSCP = y (where $x \neq y$). This mapping from one DSCP to another SHOULD be negotiated between the operators of the peer networks.

Trunk gateways SHOULD also set the appropriate DSCP on packets they generate that are destined for a DQoS endpoint. For example, this may be achieved by signalling the desired DSCP in TGCP [14] (note that the obsolete term Type of Service is used in the TGCP Recommendation). It may also be acceptable for a gateway to set the same DSCP value on all packets that it generates. If the trunk gateway cannot correctly mark packets that it generates, another device located between the gateway and the backbone (e.g. a router) SHOULD be configured to set the DSCP for IPCom media packets entering the backbone from the gateway.

It is RECOMMENDED that other IPCom elements in the media path (e.g. audio/announcement servers, anonymizers, conferencing bridges, etc.) be able to mark packets that they generate that are destined for QoS endpoints.

Within a single domain, all devices that either set or police the DSCP for IPCom media packets, or provide QoS to packets by examining their DSCP, SHOULD have consistent configuration. This may be accomplished by using COPS provisioning [13] or by other means.

It is possible to use multiple PHBs for different types of service. For example, it may be appropriate to use a different PHB for video than for voice, or it may be desired to use different PHBs for calls with tighter delay requirements due to the distance between endpoints or other factors. The mechanisms for signalling the appropriate choice of PHB in this case are described above.

7.4 PHB support by AN

Based on the criteria discussed above, one or more PHBs are selected for use in the backbone. The AN SHOULD implement all of these PHBs on its upstream links (i.e. those links connecting it to the backbone) in order to deliver appropriate QoS to packets entering the backbone. Alternatively, it may be possible to over-provision the upstream links on the AN rather than relying on Diffserv support on these links. It is important to note that over-provisioning solutions, though viable approaches to QoS, may not always be the most cost-effective or resource-efficient solutions.

7.5 Resource allocation

It is necessary to ensure that enough resources are allocated to the chosen PHBs at all network elements in the backbone. In the absence of signalling in the backbone, this is essentially a provisioning problem. COPS provisioning [13] or other means may be used to distribute provisioning information to network elements.

7.6 Admission control

Even in a statically provisioned Diffserv backbone, it is possible to perform admission control at certain points in the network. One option is to perform admission control at the AN; another is to perform admission control at the CMS. One, both, or none of these options may be appropriate. For example, if the upstream bandwidth from the AN to the backbone is large relative to the capacity of the J.112 links it serves, it may not be required to perform admission control on the AN's upstream links.

If admission control is to be performed at the AN, each AN MUST be configured with a maximum amount of bandwidth for each PHB that is to be used for IPCom media traffic on each of its upstream (non-J.112) interfaces. Each AN MUST also keep track of the amount of bandwidth that has been admitted into each PHB on each interface. When an AN receives a DQoS request to admit a call, it determines which PHB the call will use by consulting the DSCP-PHB mapping with which

it is configured, and using the DSCP provided in the GATE-SET. It MUST check to see if the amount of bandwidth available in that PHB on the outgoing interface that this call will use is sufficient to accommodate the resources required by this call. Thus, the total amount of traffic of a given PHB that will be injected into the network by any AN is bounded.

It may also be possible to perform CMS-based admission control under some circumstances. If a CMS can be provided with enough knowledge of network resources and topology, it may be able to perform admission control based on the destination of calls. For example, CMS X may know that calls which are destined to destinations handled by CMS Y must pass through a link of known capacity and can thus reject calls to that destination once the capacity is exhausted.

The admission control approaches described in this clause may have certain limitations. Notably, they may not take account of the full path through the backbone that the packets for a given call will take. Nor do they necessarily take account of the possibility of link failures affecting available capacity. Thus, there is the risk that some links will become oversubscribed. Approaches to address these issues are discussed in the following clauses.

8 Admission control for a single domain

8.1 Per-flow RSVP control plane

It is possible to use per-flow RSVP as the admission control protocol for a Diffserv cloud. An overview of this approach is provided in [4]. This clause describes the application of per-flow RSVP signalling to a Diffserv backbone in the IP-Cablecom environment. The approach described is more scalable than traditional per-flow RSVP because all classification and scheduling (i.e. all operations in the forwarding plane) are performed on Diffserv behaviour aggregates.

In order to support the capabilities described in this clause, the basic Diffserv functionality described in clause 7 MUST be provided in the AN and the backbone network. Additional requirements are presented in the following paragraphs.

To support a per-flow RSVP control plane, an AN participating in DQoS signalling MUST support the two following modes of operation:

- End-to-end RSVP mode: In this mode, the MTA signals with RSVP messages as described in clause 6/J.163 [7], and the AN MUST forward such RSVP messages towards the MTA at the far end of the call;
- Embedded signalling mode: In this mode, the embedded MTA uses J.112 signalling, and the AN MUST originate RSVP messages to the far-end MTA.

In either mode, the result is that per-flow RSVP is used between the two ANs involved in a call. It may be true end-to-end (MTA-MTA) RSVP or it may be only between the ANs acting as proxies for the MTAs; this has no effect on the backbone QoS mechanisms.

With per-flow RSVP in operation between the two ANs, a network operator has considerable flexibility as to where to place the edge of the Diffserv region. As shown in Figure 1, the edge of the Diffserv network need not necessarily be the AN, although it may be. That is, per-flow RSVP with per-flow classification and scheduling may be used from the AN to the edge of the Diffserv region; alternatively, the AN may be at the edge of the Diffserv region, in which case aggregate scheduling and classification is used on all traffic that is upstream of the AN.

For generality, we use the concept of an Edge Router (ER) as defined in [4]. This device is able to perform per-flow RSVP and to perform admission control on traffic that will enter the Diffserv network. The AN MAY perform the ER function or the function may be assigned to a router upstream of the AN, i.e. a router closer to the backbone.

In the following clauses, the behaviour of ANs, edge routers, and core routers (those Diffserv routers in the backbone which are not edge routers) are described.

8.1.1 AN behaviour

An AN may operate in one of two modes, depending on whether the MTA that it serves is using embedded signalling or RSVP signalling. Both modes **MUST** be supported. We treat each mode in turn below. In either case, the AN runs per-flow RSVP on its upstream (non-J.112) interfaces, and uses standard RSVP/Intserv procedures to perform admission control, classification and scheduling for packets sent on those interfaces.

Note that, regardless of which mode an AN operates in, it is responsible for forwarding (or originating) both PATH and RESV messages towards the far end. A bidirectional reservation is established between two ANs when a pair of RESV messages have been exchanged between them.

8.1.1.1 Embedded signalling

When using embedded signalling as defined in ITU-T Rec. J.163 [7] Annexes A and B, an AN detects the need to make a backbone reservation when a MAC-layer signalling message indicating the request to establish a new J.112 Flow arrives and when a gate has been established for the corresponding call. In this case, the AN **MUST** send a PATH message to the far-end MTA, using parameters derived from the MAC message to create the PATH as described below. It then waits for a RESV from the far-end AN or MTA. When it has received a RESV from the far-end, it knows the reservation has succeeded and **MUST** respond to the MTA with a MAC-layer signalling message indicating the success.

When an AN receives a PATH from a far-end AN or MTA, and the PATH is destined to an MTA that does not support RSVP, the AN **MUST** first verify that it has a gate established for the corresponding call. If so, it **MUST** respond with a RESV sent back to the previous hop (PHOP) contained in that PATH message. The parameters in the RESV are determined from the received PATH message.

The type and format of MAC-layer signalling messages to be used to set up J.112 Flows depends on the layer-2 protocol implemented in the cable television network. Further details on MAC signalling can be found in J.112 Annexes A, B, or C, respectively.

8.1.1.1.1 Determining RSVP PATH parameters

In order to generate an RSVP PATH message (at the receipt of a MAC-layer message indicating the need to make a backbone reservation), the AN needs to construct the session object, sender template object and the sender Tspec object. The session object consists of the protocol, destination address and destination port number. The sender template consists of the sender address and the sender port number. The mapping of the RSVP parameters to the parameters contained in the MAC message is shown in the table below.

RSVP parameter	J.112 Annex A RESC-REQ parameter	J.112 Annexes B and C DSA-REQ parameter
Session object		
Protocol Id	Session_Binding_US. Upstream_internet_protocol	Upstream packet classifier. IP protocol
Destination Address	Session_Binding_US. NIU_client_destination_IP_add	Upstream packet classifier. IP Destination Address
Destination Port	Session_Binding_US. NIU_client_destination_port	Upstream packet classifier. TCP/UDP Destination Port Start

RSVP parameter	J.112 Annex A RESC-REQ parameter	J.112 Annexes B and C DSA-REQ parameter
Sender Template object		
Source Address	Session_Binding_US. NIU_client_source_IP_add	Upstream packet classifier. IP Source Address
Source Port	Session_Binding_US. NIU_client_source_port	Upstream packet classifier. TCP/UDP Source Port Start

The sender Tspec parameters are derived from the J.112 upstream QoS parameter encodings contained in the MAC-layer signalling message requesting the establishment of the new flow. An example for mapping the MAC QoS parameters to the Tspec object to construct the RSVP PATH message is given below. For further details, refer to Annexes A, B, or C of ITU-T Rec. J.112.

The PATH message should also carry the updated Adspec object, which conveys the additional delay introduced by the AN to the RSVP routers downstream. Due to strict latency bound requirements, it is expected that hosts generating VoIP traffic would indicate their resource requirements using the guaranteed service QoS parameters as defined by the IntServ architecture. Hence, the guaranteed service block of the Adspec object should contain the appropriate C (rate-dependent component) and D (rate-independent component) terms. The value of D must take into account the fixed delay (for instance message processing delay, codec delay, etc.).

8.1.1.1.1 Constructing the Tspec object from J.112 Annex A QoS parameters

To accommodate the CBR characteristics typically exhibited by voice sources, either fixed-rate or reservation access mode may be used.

Details are for further study.

8.1.1.1.2 Constructing the Tspec object from J.112 Annexes B and C QoS parameters

Since voice sources typically exhibit CBR characteristics, it is expected that MTAs will request a unsolicited grant service (UGS) on the J.112 link. If the "service flow scheduling type" in the DSA-REQ message is set to UGS, then the sender Tspec is determined as follows. Let:

G = Grant size (bytes);

I = Grant interval (seconds).

For VoIP flows the "grants per interval" parameter would typically be set to 1 (if it is more than 1, G has to be calculated accordingly). Given above, the IntServ parameters for token bucket are:

M (maximum datagram size) = $G - \text{Ethernet overhead} - \text{ES 201 488 overhead}$;

r (bucket rate) = M/I .

The Ethernet header overhead is 18 bytes and the J.112 Annex B/C header overhead could be up to 13 bytes. Since VoIP sources exhibit CBR characteristics,

p (peak rate) = r ;

b (bucket depth) = M ; and

$m = M$.

The J.112 Annex B/C overhead includes only the MAC layer overhead (standard MAC header, BPI extended header etc.) It does not include the physical layer overhead.

If payload header suppression is being used in the upstream direction, M (as calculated above) MUST be further modified to reflect the suppressed bytes. "PHS size" parameter from the DSA-REQ MUST be used to modify M as follows:

$M' = M - 2 + \text{PHS size}$,

where two bytes constitute the J.112 Annex B/C extended header, containing the value for the PHS index. Since the grant size includes this overhead as well, it MUST be subtracted to compute M' .

The other Tspec parameters are modified accordingly:

$$r = M'/I;$$

$$p = r;$$

$$b = M';$$

$$m = M'.$$

With regard to the updated Adspec object, the advertised value of C for a UGS service would be M (or M').

8.1.1.1.2 Determining RSVP RESV parameters

When an AN receives a PATH message from the remote AN, it MUST send RESV message to reserve appropriate resources on the backbone. The RESV message MUST include the session object, flowspec and the filterspec. The session object and filterspec are derived from the PATH message. VoIP traffic should use Guaranteed Service flow specifications, which consist of a Tspec and a Rspec. The Tspec parameters are obtained from the sender Tspec object in the PATH message. The Rspec parameters are derived from the J.112 downstream QoS parameters. In a J.112 Annex B/C environment, the Rspec parameters are computed as follows:

$$R = \text{"Downstream Maximum Sustained Traffic Rate"};$$

$$S = 0.$$

The value of zero for S (slack) is the recommended value from [1] when no slack is specified.

8.1.1.2 RSVP signalling

When the MTA uses RSVP signalling as defined in clause 6/J.163 [7], and per-flow RSVP signalling is to be supported in the backbone, the AN MUST be able to forward RSVP messages into the backbone rather than simply intercepting them when received from the MTA. The AN MUST support a configurable parameter on each of its non-J.112 network interfaces that defines whether RSVP message forwarding is enabled for the interface. When that parameter has the value "enabled" on a given interface, and the AN receives a PATH message from an MTA that SHOULD be sent out over that interface according to the forwarding table of the AN, the AN MUST forward the PATH message over that interface. When forwarding such PATH message that was received from the MTA, the AN MUST remove all DQoS-specific objects (e.g. reverse Tspec, etc.) before forwarding the message on towards its destination. In this configuration, The AN SHOULD NOT proxy RESV messages back towards the MTA, but SHOULD instead wait for a RESV message to be received from the backbone and then process and forward it according to standard RSVP processing rules. Similarly, it SHOULD NOT proxy PATH messages towards the MTA, but SHOULD await a PATH message from the far end of the call instead, which it MUST process according to standard RSVP rules.

Note that the decision to forward PATH messages into the backbone rather than to perform as a proxy as described in clause 6/J.163 [7] is based on per-interface configuration. Thus, all flows of IP/Cablecom media packets traversing an interface that is configured as described above will be subject to admission control. This behaviour is desirable as it ensures that all flows entering the

network over a given interface are subject to admission control, thus permitting intelligent admission control decisions to be made¹.

8.1.2 Location of Diffserv edge

When a per-flow RSVP control plane is used across the backbone, it is not necessary for the AN to be the edge of the Diffserv cloud. Instead, the Diffserv edge function may reside in an edge router that is upstream of the AN. In this case, there exists a network between the AN and the edge router, which may be as simple as a point-to-point link (as shown in Figure 1) or may be a general topology IP network. The required QoS may be provided between the AN and the edge router either through the use of Integrated Services or by over-provisioning of the bandwidth, and the choice between these options may be made on a link-by-link basis.

8.1.3 Edge router behaviour

Whether the edge router is the AN or some other router upstream of the AN, it **MUST** participate in per-flow RSVP. In addition, an edge router (ER) has some set of interfaces that are "interior" to the Diffserv cloud and some that are "exterior". The ER is responsible for marking packets that pass from an outside interface to an interior interface with an appropriately chosen DSCP, unless it is able to trust that the DSCP was set correctly at the AN. The ER **MUST** perform admission control on all its interfaces. In order to do this on interior interfaces, each interior interface must be configured with a pool of resources available for each PHB that is to be used for IP-Cablecom media traffic. The ER performs admission control over this set of resources for each RSVP request it receives. The ER must also be able to determine which PHB and DSCP to use for a given RSVP request. This may be determined by local configuration or as a matter of policy provided from some outside source, e.g. a policy server.

An ER **MAY** perform microflow classification, policing and scheduling on its exterior interfaces but **MUST** perform aggregate classification, policing and scheduling on its interior interfaces. If the ER does not perform microflow classification and policing on flows that are passing through it into the backbone, those functions **MUST** be performed by the ANs that send traffic to the ER. Performing microflow policing at the AN may provide better scalability than doing so at the ER, as the number of flows is likely to be larger at the ER.

8.1.4 Other terminating devices (media gateways, anonymizers, announcement servers, conference bridges)

In order for per-flow RSVP signalling to operate effectively across the backbone, all devices that can terminate a media stream **SHOULD** be able to support per-flow RSVP signalling. Such devices include media gateways, anonymizers, announcement servers, conference bridges, etc. Any device terminating a IP-Cablecom media stream **SHOULD**:

- send PATH messages toward the far end(s) of the call;
- receive PATH and RESV messages from the far end(s) of the call;
- send RESV messages toward the far end(s) of the call in response to received PATH messages.

These devices derive the contents of the PATH messages from call signalling in the same way that an MTA does in normal DQoS operation. The contents of the RESV messages can be derived from PATH messages in the same manner as described in 8.1.1.

¹ An alternative approach would be to decide on a per-flow basis whether to forward PATH messages for each flow. This would raise the issue of how such a decision should be made, but more importantly it would present the risk that some subset of flows would inject traffic into the backbone without being subject to admission control, compromising the overall accuracy of admission control.

Note that, like an AN, the devices mentioned in this clause may or may not function as edge routers, in that they may be on the edge of the Diffserv cloud or not.

8.1.5 Core router behaviour

A router behaves as a core router when it receives packets on an interior interface and forwards them on an interior interface. Note that a single router may behave as an ER with regard to some flows and as a core router with regard to other flows.

A core router does not perform re-marking of the DSCP in packets that it forwards. It performs admission control over the resources allocated to the appropriate PHB for each reservation. It performs aggregate classification, policing and scheduling. Thus, the forwarding behaviour of a core router is just like any Diffserv router, even though it uses RSVP for admission control.

8.1.6 Signalling latency

The approach to bandwidth reservation described in this clause requires end-to-end RSVP messages to traverse the backbone. Clearly this may have an impact on total signalling latency and thus post-dial delay. To meet a provider's post-dial delay targets, the following techniques may be used:

- RSVP refresh reduction and reliability enhancement;
- choice of a low latency PHB and corresponding DSCP for RSVP control messages.

The same situation applies for the approach defined in the following clause.

8.1.7 Pre-emption

This clause describes mechanisms that may be used to support pre-emption of reservations (e.g. to provide resources to emergency calls in preference to previously admitted calls).

The pre-emption priority element defined for use in RSVP and COPS [6] MAY be used in the backbone. It is not expected that this object would be provided by the MTA, since end users cannot generally be trusted to determine their own pre-emption priority. However, the Gate Controller provides a Session class to the AN which MAY be used by the AN to generate a valid pre-emption priority element. In this case, the AN SHOULD use the following mapping from session class values to pre-emption priority values:

Session type	Session class value	Pre-emption priority value
Normal	0x01	32767
High priority (Emergency)	0x02	64911

This mapping SHOULD be configurable. The pre-emption priority element contains both a defending priority field and a pre-emption priority field. These SHOULD both be set to the same value.

It is also possible that RSVP-capable routers in the backbone will use COPS to outsource policy decisions. In this case the pre-emption priority element MAY be carried inside a COPS decision and its interpretation at the routers MUST be as defined in [6].

8.2 Aggregate RSVP

Aggregated RSVP [5] is a logical extension to per-flow RSVP across a Diffserv backbone. To support this functionality, the AN, edge router and core router functions described in 8.1 MUST be provided. Additional functionality is provided by aggregating and de-aggregating routers, as defined below. RSVP signalling is performed between call endpoints (either the MTAs or AN acting on behalf of MTAs) as in the preceding clause. In addition to the functionality in 8.1, aggregate RSVP defines a way in which many per-flow RSVP reservations may be combined to

form a single aggregate reservation. Two or more per-flow RSVP reservations may be aggregated when their paths pass through a common pair of routers. We refer to routers which are able to aggregate and de-aggregate reservations as aggregation routers. The behaviour of aggregating and de-aggregating routers is more formally defined in IETF RFC 3175 [5].

The aggregation routers have the responsibility of creating aggregate reservations across an aggregation region, which may be the entire Diffserv cloud or a defined aggregation region within the cloud. Each aggregate reservation represents an aggregate flow of traffic from an ingress router (or aggregator) to an egress router (the de-aggregator). Aggregate reservations may be configured statically based on the expected load from an ingress to an egress router, or they may be automatically established and re-sized as described in [5]. Each aggregate reservation carries the traffic from a number of "end-to-end" RSVP reservations that share a common ingress/egress router pair. An end-to-end reservation represents a single microflow, and signalling for such a reservation is accomplished using standard RSVP. "End-to-end" RSVP messages may be originated by the MTA or by the AN on behalf of the MTA in the case of embedded signalling, as described above. Such E2E RSVP messages are "tunneled" across the aggregation region by setting the IP protocol number in the Path message to "RSVP-E2E-IGNORE".

Note that the aggregator and de-aggregator may or may not also be edge routers as defined above. We define the relationship between these devices in 8.2.4.

8.2.1 Provisioned aggregate reservations

It is possible to provision an aggregate reservation from an ingress (aggregating) router to an egress (de-aggregating) router. This requires prior knowledge of the expected load between the routers in order to determine the size of the reservation. In this case, the ingress router sends an aggregate PATH message to the egress router, and the egress router responds with an aggregate RESV back towards the ingress. This establishes an aggregate reservation for traffic flowing from the ingress to the egress that is marked with the appropriate DSCP as identified in the aggregate RSVP messages.

Once an aggregate reservation has been established between a pair of routers, it may be treated as a logical link for the purposes of admission control. Admission control for an individual call is performed when an end-to-end RESV arrives at the egress router. Before that can happen, an E2E Path MUST be sent from the ingress to the egress. The ingress swaps the protocol ID to RSVP-E2E-IGNORE, which means that the Path is ignored by all routers between the ingress and the egress. When the egress receives the E2E Path, the PHOP (previous hop) identifies the ingress router. The egress router stores this information and then forwards the Path towards its destination.

When an E2E RESV arrives at the egress router, it determines which aggregate reservation this E2E reservation belongs to by examining the PHOP information in the Path state that matches the RESV. That PHOP is the ingress router for the appropriate aggregate reservation. The egress router MUST track the resources allocated to a particular aggregate reservation as they are consumed by admitted E2E reservations and MUST reject an E2E reservation that cannot be accommodated in the appropriate aggregate reservation.

8.2.2 Dynamic aggregate reservations

The obvious drawbacks of statically provisioning aggregate reservations is that they must be sized appropriately, and that oversizing wastes resources while undersizing will lead to excessive call blocking. These drawbacks are avoided by dynamically creating and resizing aggregate reservations in response to the arrival and departure of E2E reservations. The details of automatic creation, resizing, and removal of aggregate reservations are described in [5].

One consideration when dynamically resizing reservations is the amount of signalling overhead that may result. If the aggregate reservation is adjusted in size for every arriving or departing E2E reservation, then the signalling overhead remains equal to what it would be without RSVP aggregation, although the stored reservation state is nevertheless reduced. If excessive signalling

overhead is expected to be a problem, it is preferable to use heuristics to size the aggregate reservation, e.g. by rounding up the reserved aggregate bandwidth to something greater than the sum of the current E2E reservations.

8.2.3 Hierarchical aggregation

As defined in [5], aggregate reservations may themselves be aggregated. This may enable further reduction in the total number of reservations that need to be made through the backbone of the network, although the actual reduction clearly depends much on topology.

8.2.4 Location of aggregation points and DiffServ edge

As in 8.1.2, the DiffServ edge may be at the AN or further upstream into the backbone, and the same options apply here for provision of QoS between the AN and the DiffServ edge. Providers have considerable flexibility as to where aggregation points (aggregating and de-aggregating routers) are located. An aggregation point may coincide with the DiffServ edge (i.e. an edge router MAY perform aggregation) or it may be placed inside the DiffServ cloud. Aggregation points MUST NOT be placed outside the DiffServ cloud.

One extreme is to make the AN both the DiffServ edge router and the aggregation point. In this case the AN performs the edge router function and also performs aggregation and de-aggregation. While it may be theoretically possible to dispense with end-to-end RSVP signalling of individual flows in this configuration, end-to-end RSVP signalling provides two benefits:

- It provides a simple way to discover which aggregate reservation among many candidates is the one to which a given flow belongs.
- It provides a mechanism by which the end-points can recognize the need to dynamically create an aggregate reservation or to increase or decrease the size of an aggregate reservation.

The second benefit does not apply to statically provisioned aggregate reservations, and there are, in some cases, other ways to determine the aggregate reservation to which a single flow belongs. For example, if the aggregating and de-aggregating AN are in the same area of a network using link-state routing, the link-state database can be used to find the de-aggregator given the address of the far end MTA.

Performing aggregation at the AN leads to a potentially large number of aggregate reservations in the backbone, on the order of the square of the number of AN. If the number of calls in place between a pair of AN is typically small, then it is more useful to aggregate further into the backbone.

A given aggregation point may choose to aggregate traffic to some destinations and not to others based on a local policy (e.g. aggregate only when number of calls to that destination exceeds a configured threshold).

As in 8.1.3, microflow policing MUST be performed before a flow's packets enter the DiffServ cloud. This function may be performed by the AN or the edge router.

8.3 Bandwidth broker

The notion of a bandwidth broker (see Figure 2) is introduced in [8] and has been the subject of considerable research. A bandwidth broker is a centralized admission control agent from which requests for bandwidth can be made. Such requests may be made by hosts, by other brokers in neighboring domains, or by edge routers. In the IPCablecom environment, it would be possible for each AN or CMS to make requests for bandwidth from a bandwidth broker that was responsible for managing access to the bandwidth for a domain. These requests would also specify the PHB for which the request is being made. The bandwidth broker for each domain is then responsible for making requests for bandwidth from neighboring domains.

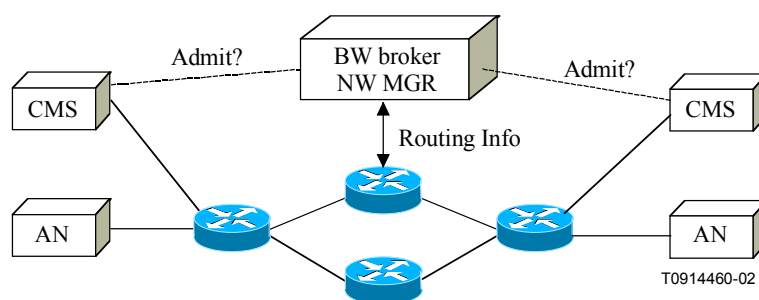


Figure 2/J.174 – Bandwidth broker example

There is considerable flexibility in the admission control algorithm and mechanisms that the bandwidth broker may use. Each broker must reject any request for bandwidth for a given PHB that would result in over-commitment of resources and degradation of the quality of calls already in progress. In order to perform this admission control function, a bandwidth broker may simply bound the total amount of traffic that is allowed to enter the domain without regard to the paths that calls will traverse. In this case, the network operator may make some statistical assumptions about the distribution of calls (e.g. that it is very unlikely that all calls will converge on a single link) in order to determine the amount of bandwidth that may safely be granted. A more conservative approach would be to assume the worst case in which all calls converge on the most resource constrained link can happen, and to use the capacity of that link for the request PHB as the bound on admitted bandwidth requests.

A more sophisticated approach to admission control would rely on the bandwidth broker having some understanding of the network topology and the route to be taken by a call. A bandwidth broker could be configured with knowledge of the network topology (perhaps limited to the location of the most resource-constrained links), or it could dynamically learn the topology, e.g. by listening to link-state routing advertisements, enhanced with resource information. The requests for bandwidth made to the broker in this case must include enough information about the destination of the call to allow the broker to determine which resource-constrained links this call will traverse and thus whether the call can safely be admitted.

As with aggregated RSVP, it is not strictly necessary for per-call signalling to take place – it may be possible for an AN to aggregate requests for calls with similar destinations. Note, however, that this would require some topology knowledge in the AN.

There does not currently exist a standard protocol for communication with or between bandwidth brokers.

9 Admission control over multiple domains

It is expected that individual IPCablecom domain networks will have their own policies and operational procedures. It is also expected that IPCablecom network operators may use a variety of IP transport providers to carry their IPCablecom traffic, each of which may employ different network topologies. It is therefore difficult to assume that consistent QoS mechanisms will be available end-to-end for calls that cross the backbones of multiple providers. As described in [4], it is possible to use RSVP end-to-end without requiring that all intervening domains be RSVP-aware. For example, one domain might use a pure provisioned Diffserv model, another might use RSVP aggregation, and another might use per-flow RSVP. One observation that can be made is that there is no reason per-flow or aggregate RSVP reservations cannot traverse domain boundaries if two adjacent domains agree to honor each others' RSVP requests. In such an environment, stronger assurances may be obtained than would be possible if some domains do not support RSVP. In addition, the effect of aggregation on scalability may be improved if aggregate reservations are able

to traverse domain boundaries, as this avoids the need to de-aggregate the RSVP requests at the border router.

It is also possible for different providers to choose widely varying technological approaches for providing QoS in their backbones. For example, one provider may choose to implement its backbone using ATM, and RSVP reservations (individual or aggregated) may be satisfied by establishing ATM virtual circuits with appropriate QoS characteristics. Other providers may use Synchronous Digital Hierarchy links to directly connect routers. Providers have similar flexibility in deciding whether or not to use MPLS as discussed above.

10 Use of MPLS

NOTE – The material presented in this clause is for informational purposes only.

Multiprotocol Label Switching (MPLS) [12] MAY be used in the backbone, with label switched paths (LSPs) being used to represent aggregate reservations or aggregate traffic flows. This offers the following potential benefits:

- 1) ability to perform traffic engineering more precisely than without MPLS;
- 2) recovery mechanisms around link and node failures;
- 3) constraint-based routing of aggregate reservations;
- 4) consistent routing of aggregate control messages and data.

The first benefit applies to all of the approaches described in the preceding clauses. By traffic engineering we mean the ability to control the paths taken by aggregate flows of traffic, with the general goal of avoiding over- or under-utilization of links. MPLS traffic engineering and its benefits are described in [11].

As described in [15], MPLS also provides facilities to protect against the failure of links or nodes in a network. For example, backup paths can be pre-established to bypass, and thus protect against the failure of, a link or node. By routing packets onto a backup LSP from a node upstream of the point of failure, it is possible to avoid the delay associated with waiting for IP routing to re-converge after a failure. Thus, the period of time for which forwarding of packets is interrupted due to either link failure, node failure, or packet loss arising from inconsistent routing, can be significantly reduced.

Constraint-based routing of aggregate reservations enables paths to be selected based on their ability to satisfy constraints, notably the availability of sufficient bandwidth to accommodate a particular reservation request, as described in [11]. This will typically allow more reservations to be established than would be possible if all reservation requests followed the shortest path as determined by conventional IP routing.

The fourth benefit, as discussed in [5], primarily applies when traffic is split across equal cost paths, introducing the risk that an aggregate PATH message would take one path while the data requiring a reservation would take another. This issue is avoided if the data is "tunneled" from ingress to egress, and MPLS provides a suitable tunneling technology.

An additional benefit of MPLS is that it can be deployed incrementally on a node-by node basis via software upgrades. This is beneficial in that existing routing and QoS mechanisms can be preserved and supported during a phased MPLS roll-out.

The choice of whether to run MPLS in any domain can be made independently from other domains and depends on whether the provider needs or wishes to address the three issues listed above.

Note that MPLS may be used with any of the approaches described in clauses 7 and 8. In a purely (non-signalled) Diffserv backbone, the primary benefits would be traffic engineering and fast reroute. Constraint-based routing of reservations is useful for either of the approaches in 8.1 and 8.2, while the consistent routing of control and data messages is only significant for aggregate RSVP (see 8.2).

11 Queuing and filtering

11.1 Queuing

Voice quality over IPNetwork networks may not only be impaired by bandwidth restrictions. Latency, jitter and packet loss are also transmission parameters that need to be observed. Reduction of delay and jitter is mainly achieved by a proper management of queues in routers. As such, the choice of queuing technologies may be as important as the choice of QoS mechanisms such as Diffserv, per-flow RSVP, and RSVP aggregation are used. Carefully chosen and properly configured queuing methods can give low jitter and low latency to the specified traffic.

Priority Queuing (PQ) for example can give a higher priority to IPNetwork media traffic over other kinds of traffic. Prioritization allows other packets, besides IPNetwork media traffic, to use the queue as long as the QoS for IPNetwork traffic is maintained. However, with PQ the fair treatment of low priority traffic has to be ensured.

Weighted Fair Queuing (WFQ) can distribute the available bandwidth evenly among traffic flows. Even in congested situations, a constant throughput is maintained. However, as this method cannot give absolute priority to the IPNetwork media traffic, quality degradation is to be expected in congested routers.

Several other methods that generate queues dynamically or use a multitude of queues may also be implemented. The applicability of those methods for IPNetwork needs to be evaluated against their ability to maintain the specified QoS for the media traffic. A careful configuration of the queuing method will often be needed.

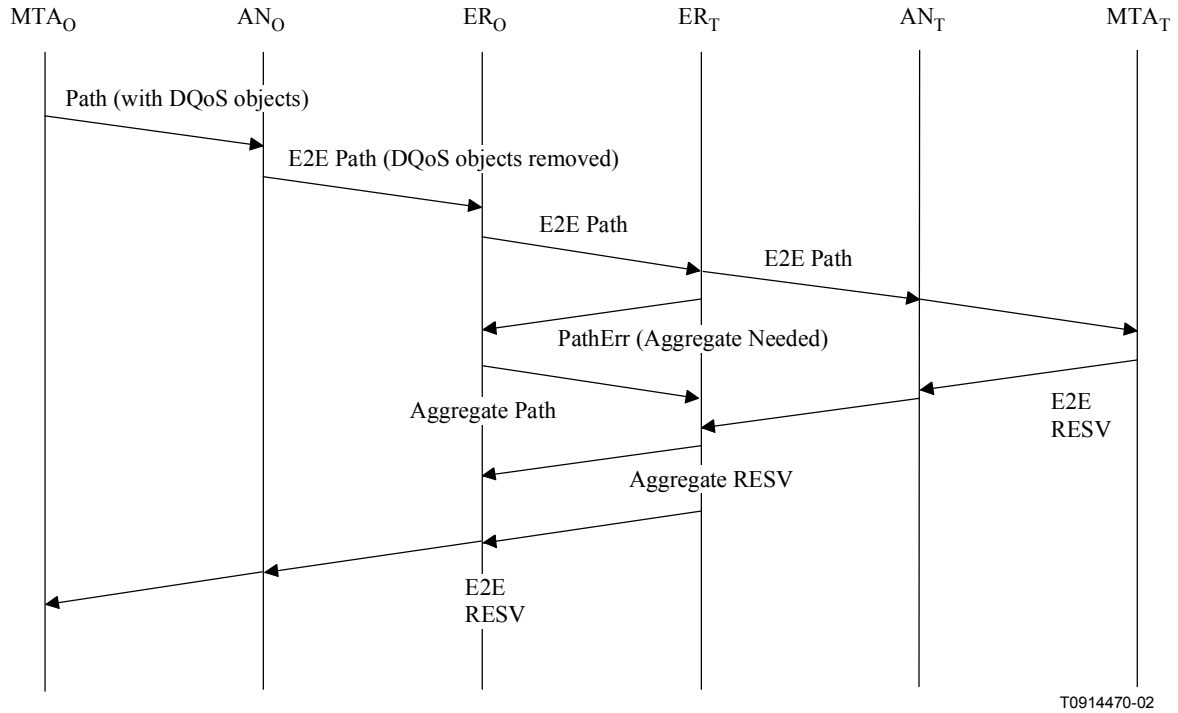
11.2 Filtering

Some routers employ filtering functions to limit traffic for reasons of network security. Filtering functions may consume resources, such as CPU power and memory, which may impair router performance and, therefore, cause degradation of QoS parameters such as packet loss, jitter and latency. Routers that use packet filtering should have enough resources to handle all the traffic they may encounter.

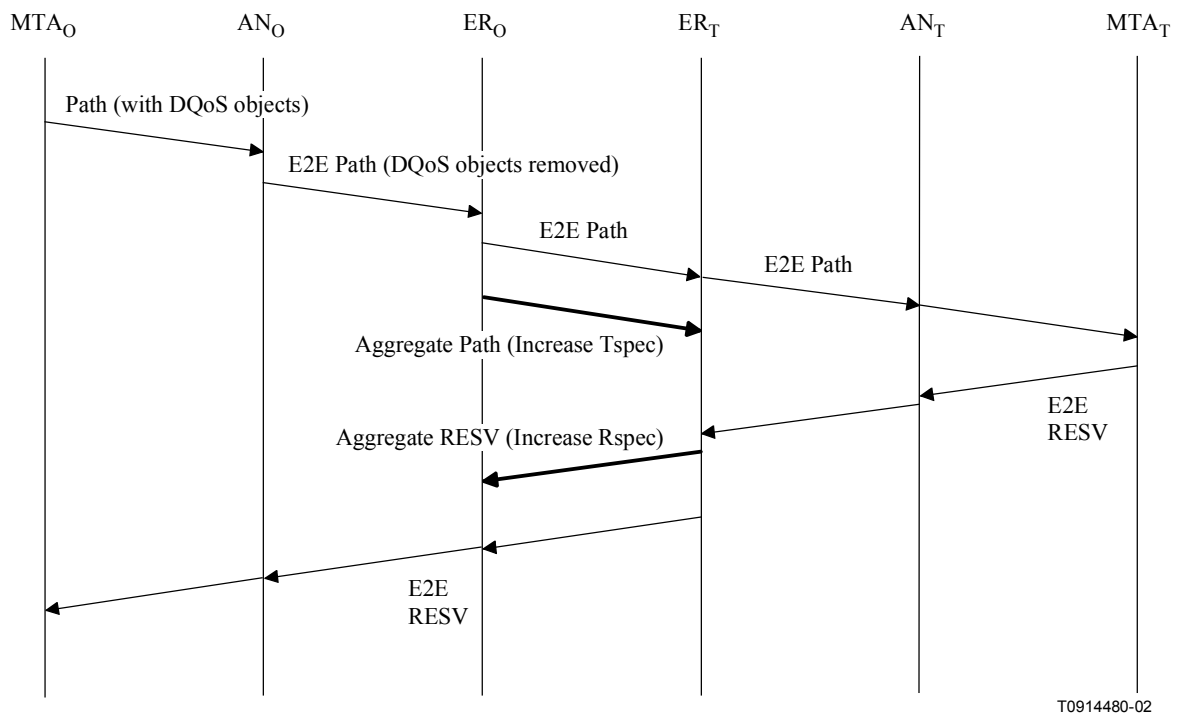
Appendix I

Call flow examples

In the examples, both MTAs use DQoS RSVP signalling and RSVP aggregation performed by ERs; only unidirectional message exchange is shown for clarity.



The following example shows increase in size of an existing aggregate reservation in response to a new E2E reservation.



Appendix II

Bibliography

- [1] ITU-T Recommendation Y.1541 (2002), *Network performance objectives for IP-based services*.
- [2] IETF RFC 2205 (1997), *Resource ReSerVation Protocol (RSVP) – Version 1 Functional Specification*.
- [3] IETF RFC 2210 (1997), *The Use of RSVP with IETF Integrated Services*.
- [4] IETF RFC 2211 (1997), *Specification of the Controlled-Load Network Element Service*.
- [5] IETF RFC 2408 (1998), *Internet Security Association and Key Management Protocol (ISAKMP)*.
- [6] IETF RFC 2409 (1998), *The Internet Key Exchange (IKE)*.
- [7] IETF RFC 2747 (2000), *RSVP Cryptographic Authentication*.
- [8] IETF RFC 2748 (2000), *The COPS (Common Open Policy Service) Protocol*.
- [9] IETF RFC 2753 (2000), *A Framework for Policy-based Admission Control*.
- [10] AWDUCHE [D.] et al.: Extensions to RSVP for LSP Tunnels, *draft-ietf-mpls-rsvp-lsp-tunnel-09.txt*, August 2001.

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of 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	Cable networks and 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 and Internet protocol aspects
Series Z	Languages and general software aspects for telecommunication systems