ITU Telecommunication Standardization Sector Document AVD-2618

Study Group 16 Version 1.0

Q. 2,3,4,5,21,22,24,25 and 29/16 Rapporteur Meeting Feb 2005

Melbourne, 28 February - 4 March 2005

Question(s): F/16 and 2/16

Source*: Subha Dhesikan, Cisco Systems

Seong-Ho Jeong, HUFS/Samsung

Title: New Text for Draft New Recommendation H.323 Annex N

Purpose: Proposed New Text for H.323 Annex N

End-to-End Quality of Service (QoS) and Service Priority Control Signalling in H.323 Systems

*Contact:	Seong-Ho Jeong	Tel: +82-31-330-4642
	HUFS/Samsung	Fax:+82-31-333-4256
	Korea	E-mail: shjeong@hufs.ac.kr
	Subha Dhesikan	Tel: +001-408-902-3351
	Cisco Systems	Fax: +001-408-902-3351
	USA	E-mail: sdhesika@cisco.com

Table of Contents

1 Introduction

This draft defines the H.323 QoS and service priority mechanisms for authorising, signalling and controlling end-to-end QoS and service priority in calls involving multiple network operator domains, multiple service domains, and heterogeneous transport and QoS mechanisms (e.g., mixed IP, ATM, and MPLS environments.)

Supporting the end-to-end quality of service (QoS) and service priority mechanism defined in this draft is optional in a H.323 system. However, the H.323 entities shall not abruptly fail on receipt of a QoS object/message but it is instead preferred that they disregard the received objects.

It is recommended that the QoS mechanisms defined in Appendix 2 of the H.323 document and the mechanism here in this draft is combined and published separately from the H.323 specification.

2 Author's Notes

This section contains a brief summary of what is new in this version of the document. It also contains a list of points for which comments are requested. Please forward any thoughts you have on these points.

This version is built upon the current version of Annex N. One significant addition is the authorization function. Authorization and authentication is essential to ensure that only legitimate requests are allowed access to network and application resources. Therefore, it is important not only in the H.323 application space but also in the network space. The authorization function is discussed in section

Another significant modification is that the polling mechanism is included in addition to the existing push mechanism that has been defined between the H.323 applications and network domains. This has been defined in section 6.5 and 7.4.

This draft also includes a strength indication to the QoS requests. In some domains, it may be preferred that a call is not established unless the required QoS is available. However, in other domains, the calls may be allowed to be established whether the desired QoS is available or not. Annex N includes a new parameter called General Bearer Descriptor. This document has incorporated the use of this parameter and expanded the QoS signalling to use this parameter. In addition, the following points are also made:

- It is recommended that appendix II be folded within this document and that the same techniques and parameters be defined to enable QoS whether single or multiple domains.
- Most of this document pertains to gatekeeper-routed call signalling as this option is considered to be more prevalent when multiple domains are involved. However, it is expected that this document will apply just as well to direct-routed call signalling and variations thereof. Future versions will include coverage for all such options.
- The author would like to initiate a discussion on the need for "gatekeeper controlled services" defined in the current version Annex N. Instead of endpoint-gatekeeper negotiating the type of QoS services, such as guaranteed or controlled load, a negotiation of the QoS strength, such as desired versus required, may be better suited.

- While the objects and techniques defined here in verified for transport-QoS mechanisms for IP (DiffServ and RSVP) and MPLS. The author requests assistance to ensure the same for ATM.
- The author intends to check on H.460.14 and reconcile any overlaps with priority. Any assistance in this regard is welcome.

3 Scope

When H.323 signalling is deployed between endpoints in a single domain or domains employing homogeneous transport QoS mechanisms and policies, the techniques described in appendix II may be sufficient to achieve bearer QoS. Whereas, in environments, where a call has to traverse multiple network domains, the transport systems in the network domains may not be homogeneous and therefore may support different QoS mechanisms and policies. This annex is targeted towards such environments. This document specifies techniques to enable a call to obtain the necessary QoS and service priority for its streams in such environments. The following points further bounds the scope of this document:

- This annex defines the mechanisms (control parameters, message formats and procedures) between H.323 functional entities that may be used for signalling and control of end-to-end QoS and service priority in H.323 systems. Mechanisms required for signalling between H.323 entities and network domains, and QoS signalling between the entities in the network domain are not covered in this document.
- The QoS and service priority mechanisms described in this annex are also independent of transport QoS mechanisms that occurs between the entities in the network domain (e.g., Differentiated Services (Diffserv) or Integrated Services (IntServ) /RSVP or ATM QoS mechanisms).
- A media independent generic syntax is defined in this document for characterising bearer QoS and service priority. The syntax may also be used by H.323 functional entities and in Service Level agreements and specifications. However, that is outside the scope of this document.
- A monitoring function or H.323 MIB to monitor the delivered QoS is considered important but is currently outside the scope of this document.

4 Requirements

The H.323 QoS and service priority mechanisms shall meet the following functional requirements:

4.1 General

- **Quantitative Guarantees:** The mechanisms described here shall facilitate guaranteed end-to-end QoS as well as best effort operation. A framework that only improves best effort communications or provides relative guarantees (e.g., voice gets better treatment than other services) is considered insufficient in all circumstances.
- **Service Priority:** It should be possible to indicate priority of service within H.323 systems. Such priority may be specified between service providers, and between service

providers and end users. Media flows categorised as high priority shall take precedence over those categorised with a lower priority in the allocation of a transport resource.

- **Network Interoperability:** The mechanisms shall allow operation over a concatenation of separate network operator domains.
- **Delay:** The mechanism should not contribute adversely to call or media stream set-up times.
- **Security:** The mechanisms should be independent of and compatible with H.323 security mechanisms.

4.2 Endpoints

- **QoS Class Discovery:** Endpoints shall be able to discover the H.323 QoS service classes and priority classes supported by the H.323 system. In the case of static SLAs, the default QoS and service priority levels may be discovered by using the pre-call setup signalling messages.
- **Endpoint QoS Selection:** Endpoints shall be able to indicate desired QoS and service priority preferences for both transmit and receive media streams during both the pre-call setup and the call setup phases. The mechanisms defined should support request, confirmation (by both parties and/or network entities) and negotiation of QoS and priority classes.
- **Default and Per-call Control:** It shall be possible to signal QoS and service priority preferences on a per call basis as well as to register the end point’s QoS and service priority preferences with the service provider that can then be used as defaults. It shall also be possible to operate using the default QoS and service priority preference settings and optionally override on a per-call basis.
- **Endpoint’s Characteristics:** Each endpoint shall be able to indicate its QoS performance and characteristics to the gatekeeper. This is required because the endpoint itself contributes to the end-to-end QoS performance.

4.3 Gatekeepers

- **Users Registered QoS Profile:** It shall be possible to register the users’ QoS and service priority profile (the QoS levels and service priority that the user is entitled to request and receive) with the service provider.
- **Mobility:** It shall be possible for mobile terminals to initiate calls requesting specific QoS and service priority preferences in accordance with their registered profile. It shall also be possible for mobile terminals to register their QoS characteristics and capabilities with service providers and receive calls in accordance with these registered characteristics.
- **Gatekeeper QoS Role**. The gatekeeper shall support mechanisms to determine if an end-point's QoS and service priority preferences are achievable.
- **Inter-Gatekeeper QoS Communication**. Gatekeeper QoS mechanisms shall be extensible to inter-domain environments. Gatekeepers shall be able to communicate QoS and Service priority preference and characteristics between service domains.

- **Static and Dynamic Inter-Domain Operation.** Gatekeeper(s) may signal QoS and service priority preferences between service domains on a per-call basis or operate on the basis of service level agreements using static QoS provisioning.
- **Transport Network Interfacing.** Gatekeepers shall be able to communicate with transport network QoS entities to establish the ability of the transport network to support QoS requirements, to instruct the transport networks to establish specific calls and media streams across the network with specified QoS requirements, and to provide appropriate authorisations. However, the scope of this annex shall be limited to extending H.225.0 and H.245 signalling messages and the mapping of the application layer H.323 QoS classes (or service classes) to network layer QoS parameters. The definition of a specific transport layer interface is out of scope for this document.
- **Authorisation**. The gatekeeper shall determine whether an endpoint's QoS and service priority preferences are within the endpoints' QoS and service priority profile and grant authorisation for the establishment of a call or media stream. The definition of authorization mechanisms within and between gatekeepers and policy servers and/or protocols to support such mechanisms is outside the scope of H.323 Annex N.

4.4 Transport Mechanisms

- **Operation over Heterogeneous Transport Networks.** It shall be possible for the mechanisms to operate over packet networks employing different transport technologies and/or different transport QoS mechanisms. The mapping of the application layer H.323 QoS classes (or service classes) to network layer QoS parameters is intended to provide a common basis for QoS control across all heterogeneous transport network layer QoS.
- **Utilizing Existing Mechanisms.** The H.323 mechanisms shall be compatible with existing transport QoS mechanisms and provide co-ordination between them via mapping of the application layer H.323 QoS classes (or service classes) to network layer QoS parameters. As an alternative to dynamic transport QoS mechanisms, statically configured or provisioned transport networks may be used to provide QoS guarantees.
- **Service Interoperability.** The mechanisms shall allow non H.323 data services to share the same transport infrastructure.

5 Architecture

The architecture described in this section is based on H.360 (An Architecture for End-to-End QoS Control and Signalling). H.360 provides reference architecture for defining and analysing mechanisms and procedures for achieving end-to-end QoS and service priority control.

5.1 Functional Planes

To achieve end-to-end QoS in H.323 systems, the H.323 QoS mechanisms operating at the application level must interoperate with the QoS mechanisms operating in the transport network (e.g., RSVP, DiffServ etc). Furthermore, network management functions may also be involved in controlling and managing QoS. Figure 1 illustrates the relationship between the H.323 Application Plane, Packet Based Network Plane, Circuit Switched Plane (e.g., PSTN), and Management Plane for the general case where the end-to-end system is made up of both a PSTN portion and an H.323 portion.

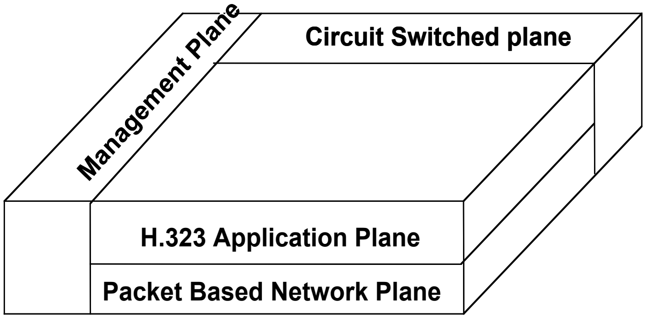

Figure 1: Functional Planes

5.1.1 H.323 Application Plane

Within this plane, QoS parameters specific to the H.323 application (e.g. QoS class, speech quality, end-to-end delay) are requested, authorised, signalled, controlled and accounted.

5.1.2 Packet-Based Network Plane

Within this plane, general non-application specific parameters effecting QoS, (e.g., end-to-end delay, delay jitter, packet loss and bandwidth) must be controlled and accounted to achieve the QoS requirements requested by the H.323 application.

5.2 The H.323 System

The H.323 System, as defined in this document, includes the functionality of both the H.323 application plane and the transport plane. The H.323 application plane is, in general, made up of a number of separate H.323 service domains, each representing the domain of control of an H.323 end-user or H.323 Service Provider. Examples of H.323 entities within the service domain are gatekeepers, gateways, H.323 endpoints etc.

The transport plane includes a number of separate network operator domains. The network domains consist solely of transport related functionality and that includes IP routers and switches, firewalls, etc. Each network domain may have its own QoS policies and/or differ from other domains in terms of administrative control (e.g., network operator), QoS mechanisms (RSVP/IntServ, DiffServ, and MPLS etc), access, metering, addressing schemes (global, local) and transport protocol (IPv4, IPv6), etc. The general H.323 system deployment is illustrated in Figure 2.

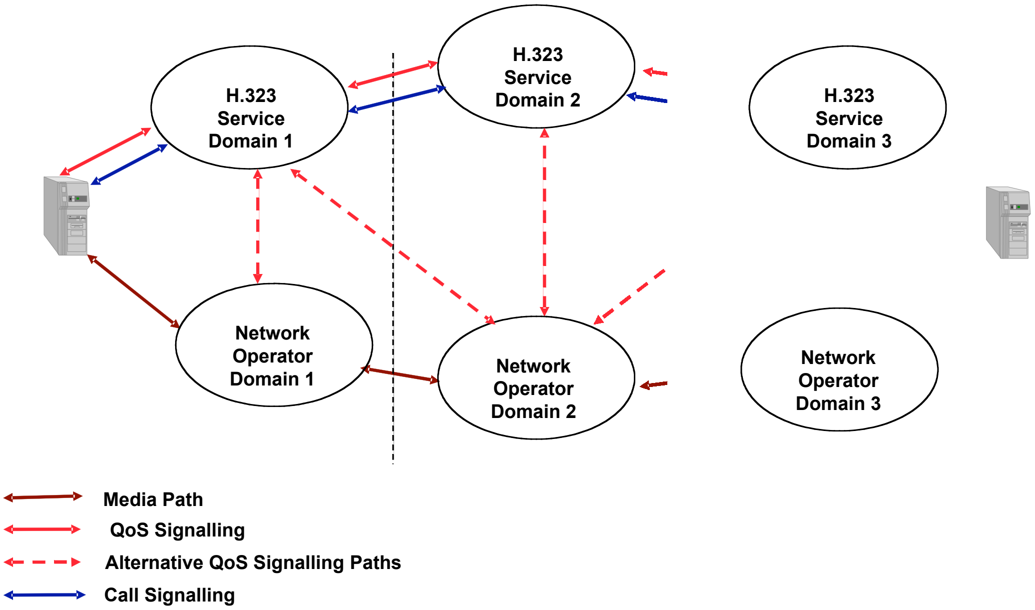
H.323 Service Domain 1
H.323 Service Domain 2
H.323 Service Domain 3
Network Operator Domain 1
Network Operator Domain 2
Network Operator Domain 3
Media Path
QoS Signalling
Alternative QoS Signalling Paths
Call Signalling

Figure 2: H.323 call involving multiple network operator domains

[Author: The figure and the associated text seem to only cover gatekeeper-routed call signalling. A subsequent version needs to address the other call signalling methods]

In Figure 2, call signalling, H.323 QoS signalling and media flows are shown separately. QoS signalling within the transport network are not shown as in the above figure and is outside the scope of this draft. The H.323-QoS signalling mechanisms involve signalling of QoS parameters between H.323 domains, and between H.323 domains and end-points. This QoS signalling is the focus of this draft.

Each H.323 service domain may also interact with one or more network operator domains involved in transporting the bearer. To specify and request the required bearer QoS, QoS signalling will take place between the controlling H.323 service domain and the network operator domain(s) providing the transport service. This QoS signalling, shown in figure 2 above in dotted lines, is mostly outside the scope of this draft. However, this is area of signalling is briefly discussed in section 5.2.2.

5.2.1 Functional Entities

Taking a closer look at a single H.323 service domain and its associate network domain, the following functional entities can be identified. They are:

- **QoS Service Manager (QoSM).** A functional entity that mediates requests for end-to-end QoS in accordance with the policy determined by the QoSPE. It communicates with other QoSMs and with TRMs to determine, establish and control the offered QoS.

- **QoS Policy Entity (QoSPE).** A functional entity that manages H.323 application policies and provides authorization of permitted and default QoS levels. It receives requests from and issues responses to QoSMs to establish the authorized end-to-end QoS levels.
- **Transport Resource Manager (RM).** A functional entity that applies a set of policies and procedures to a set of transport resources to ensure that these are allocated to enable QoS guarantees across the domain of control of the TRM
- **Transport Functionality (TF).** A functional entity representing the collection of transport resources within a Transport Domain which are capable of QoS control.
- NPE:

The relationship between these functional entities is shown in Figure 3.

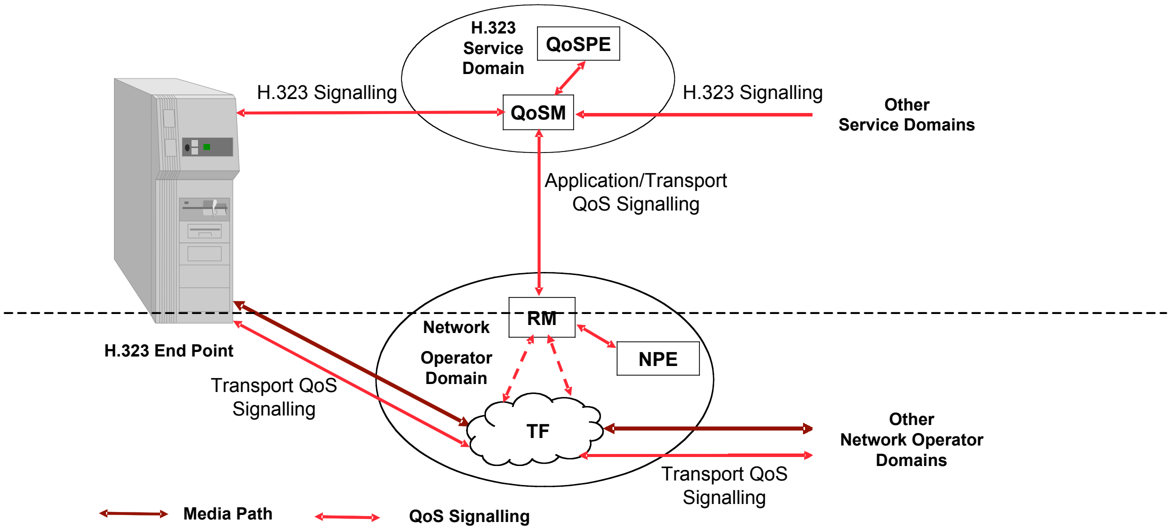

Figure 3: Relationship between QoS Functional Entities

The QoSM will normally be functionality within an H.323 gatekeeper while the QoSPE may be functionally within a back-end policy server.

When the QoS mechanisms and policies are homogenous, then the approach described in appendix II can be applied. However, in reality, a call between H.323 endpoint may traverse multiple domains employing different QoS mechanisms and policies. This motivates the need for the approach described here, which requires that some aspects of resource reservation and authorization be performed by signalling between the QoSMs and the relevant RMs.

The signalling protocol used between QoSMs and RMs, QoSMs and QoSPEs, and RMs and NPEs are not within the scope of this document.

5.2.2 Interactions between the Application and the Transport Planes

Though the specifics of the interaction between the application plane and the transport plans is outside the scope of this draft, a brief discussion helps to bind and justify the rest of the QoS material. The main points to consider is what kind of QoS signalling occur in the network, whether there are sufficient to carry the QoS requests, how is the network device/interface identified and how does the authorization occur.

One option assumes that the H.323 entities have the necessary network awareness. With this awareness, the H.323 entity can identify the network device/interface that is going to service the bearer stream. Therefore, it can push the QoS request and authorization for the bearer stream to that network device/interface. This is an attractive mechanism but may not work in all environments. This may be more suitable in small or homogenous network but in other large or heterogeneous networks, the requirement of network awareness creates a scaling problem.

Another option is the signal the QoS request via a path-coupled mechanism (like RSVP). The network device/interface that receives the request approaches the H.323 entity for authentication and authorization of the request. This is scalable and works in complex topologies but requires the use of an explicit signalling protocol (RSVP) and also requires the maintenance of state in the network devices that renders it less attractive.

The third option may be a mix of the two, where a media relay is used. The media relay participates in both the call signalling and remains in the path of the bearer stream. The media relay knows the bearer stream to make the necessary request to the underlying transport system and also is in a position to authorize the request as a valid one. This option too has scaling concerns.

There are no clear winners among the various options not only because each of the options has its own share of cons but because the options depend on the network domain and the QoS support that it extends. For example: In a pure DiffServ network, one does not expect to use RSVP. The objective that is maintained in this draft is to enable all the 3 options so that each H.323 subsystem can choose and deploy any one of the options. The options may be used individually or together in a single call.

6 QoS Negotiation

To enable end-to-end QoS, the H.323 entities must agree to participate in the QoS methods and agree upon the QoS parameters. These agreed-upon QoS parameters are then translated by H.323 entities along the path to QoS requests to each of their associated network domains. A new parameter, called Generic Bearer Descriptor, is defined to enable the QoS negotiation among the various H.323 entities.

6.1 Generic Bearer Descriptor (GBD)

QoS parameters may be signalled within the H.323 system by use of a Generic Bearer Descriptor (GBD). The descriptor contains 4 elements which provide the QoS requirements, description, priority and any necessary authorization needed for the bearer stream. The 4 elements are:

- Service priority parameter: Indicates the priority of the stream
- QoS Descriptor: Provides the QoS requirements for the stream.
- Traffic Descriptor: Provides the traffic profile of the stream

- Authorization elements: Policy elements that authorize the request.

These elements are explained further in the section below. The GBD may also be used between the H.323 domains and the network operator domains but this is outside the scope of this draft.

6.2 Service Priority

[Author: Need to reconcile any overlaps with H.460.14]

The service priority parameter shall be used to signal the priority of service to be provided to a traffic stream within an H.323 system. This priority parameter may be signalled between service providers, and between service providers and end users. Media flows categorised as high priority shall take precedence over those categorised lower in priority with respect to the allocation of transport resource. The initiating service provider shall determine the priority to be assigned to the media stream and signal this to other service providers or endpoints involved in the call.

There are 3 possibilities with respect to the service priority. They are:

- No service priority is required for the stream. This is indicated by the absence of the service priority parameter.
- A service priority is required but is not specified as they are available from other sources such as static configurations and service level agreements
- A service priority is required and is signalled via the service priority parameter.

To allow for the above possibilities (2 & 3 above), the following format of the service priority parameter r is provided:

- **ServicePrioritySignalled (boolean)**
  This parameter shall specify whether service priority is to be signalled using the **Service PriorityValue** parameter. A false value indicates that the service priority is based on a value determined by a priori agreement between business entities.
- **ServicePriorityValue (enumeration)**
  This parameter shall be used to specify the service priority of the bearer.

[*Author: It is suggested that a range be provided instead]*

6.3 QoS Descriptor

The QoS descriptor contains the QoS requirements for the bearer stream. It includes a type of request followed by required QoS values. The **qosType** is used by the H.323 system to decide whether a call is to be continued or failed based on the QoS success and failures. The **qosValue** are described in terms of maximum end-to-end delay, maximum end-to-end delay variation, and maximum mean packet loss. It is necessary to signal this information end-to-end as it allows the intermediary H.323 entities to negotiate these QoS requirements with their respective network domains. The QoS descriptor is described in detail below:

QoSType:

The QoS type indicates the strength of the QoS request. There are 3 possibilities with respect to the QoS type. They are:

- **No QoS** is required for the stream: This is indicated by the absence of the QoS descriptor. For example: If best effort service is sufficient for the stream, then no QoS descriptor parameter need to be signalled.
- **Desired:** This indicates that QoS is desirable but not mandatory for the call. This means that the QoS request must be attempted but the call can continue even if the desired QoS is not granted.
- **Required:** This indicates that QoS is required and the call cannot continue if the required QoS for the stream is unavailable.

QoSValue:

The next object is the QoS value. The **qosValue** may be left unspecified if they are to be derived from other sources such as static configurations and service level agreements. They are to be signalled via the **qosValue** parameter if they are to be specified. The **qosValue** contain the following information:

[Author: One suggestion is to replace the exchange of this following information with the QoS class information. One expects that every QoS-aware H.323 entity knows the mapping of each QoS class to the following parameters is known to every H.323 entity. Therefore the H.323 entity can translate the QoS class to the transport level parameters as it is neede.]

- Maximum end-to-end delay
  - MaxDelayClass (enumertion)
    This parameter shall specify a number representing an entry in a list of possible maximum delay values and based on a priori agreement between business entities. The values of maximum delay to be included in this list are for further study.
  - MaxDelayValue (numeric)
    This parameter shall be used to specify the numerical value of the maximum delay.
- Maximum end-to-end delay variation
  - MaxDelayVariationClass (enumeration)
    This parameter shall specify a number representing an entry in a list of possible maximum delay variation values and based on a priori agreement between business entities. The values of maximum delay variation to be included in this list are for further study.
  - MaxDelayVariationValue (numeric)
    This parameter shall be used to specify the numerical value of the maximum delay variation.
- Maximum Mean Packet Loss
  - MaxMeanPacketLossClass (enumeration)
    This parameter shall specify a number representing an entry in a list of possible maximum mean packet loss values and based on a priori agreement between business entities. The values of maximum mean packet loss to be included in this list are for further study.
  - MaxMeanPacketLossValue (numeric)
    This parameter shall be used to specify the numerical value of the maximum mean

packet loss. [Note: The use of a parameter to indicate the maximum permitted levels of packet loss in a burst of specified duration is for further study]

6.4 Traffic Descriptor

The traffic profile of the media stream is required for signalling to the transport plane. The transport plane utilizes such information for resource management and admission control purposes. The agreed-upon QoS levels for a stream are guaranteed only if the flow remains conformant to the traffic profile provided. The parameters of a traffic descriptor are as follows:

- **PeakBitRate (numeric)**
  This is defined as the maximum rate of the media stream at which the transport domain is required to sustain QoS guarantees. (The header overhead from RTP or equivalent framing shall be included in this figure).
- **PacketSize (numeric)**
  This is defined as the maximum size of the media packets including the RTP header overhead.

For constant bit rate (CBR) media flows, the peak bit rate is sufficient to enable optimum resource utilisation in the transport plane. However variable bit rate (VBR) media stream may require further parameters such as the average bit rate of the flow, burst etc.

[Author: The extension of the traffic descriptor to characterise VBR media flows is to be completed. The parameters believed required are average packet rate and burst.]

6.5 Authorization Elements

The authorization elements considered are:

- Token: A token may be described as a hash/MD5 of various pieces of information. The information includes a secret key, bandwidth required (transport descriptor), source and destination information of the bearer stream, etc.
- This includes the identity of the bearer stream, the traffic descriptor and a secret token.
- Requestor: This information is required so that the network entity can poll the requestor for authorization.

[Author: The various authorization elements need more analysis. This should be in-line with the general security mechanisms defined for H.323.]

7 H.323 QoS & Service Priority Control Procedures

The QoS and service priority procedures are included in various stages of call-setup. The procedures in each stage are discussed in detail below:

7.1 Pre-Call Setup Procedures

This is the discovery phase of the QoS and service priority setup. The following steps are involved:

- **System QoS Discovery:** First of all, the endpoints need to discover the QoS and service priority classes supported by the H.323 system and any defaults provided as well.
- **Default Class Selection:** The next step is the endpoint selection of a default H.323 QoS and service priority class applicable to all calls or media streams established from that end point.
- **Transport-QoS capabilities negotiation:** In this step, the endpoint indicates its support of the transport QoS mechanisms to the gatekeeper. The gatekeeper either accepts or rejects the capabilities and indicates its choice. This is done during the RAS exchange.
- **Gatekeeper user profile discovery:** Discovery by a gatekeeper of the profile of a visiting user to the service domain controlled by the gatekeeper.
- **Gatekeeper to Gatekeeper Service Class Discovery:** Gatekeeper discovery of the H.323 QoS and priority classes supported by another gatekeeper or the default QoS and priority levels provided by the system.

7.1.1 Endpoint QoS Capabilities Registration

An endpoint should advertise its transport-QoS capabilities for enabling end-to-end QoS to its gatekeeper. Currently, the capabilities include the ability of the endpoint to participate in RSVP-based admission control and to mark the media packets with an appropriate DSCP value. This capability is signalled during the RAS exchange using the **transportQoS** field of the RRQ, ARQ, or LRQ message. If sent in the RRQ or LRQ message, the capabilities expressed in the **transportQoS** field apply to all calls made by the endpoint, unless the endpoint overrides the capability by specifying a **transportQoS** field in an ARQ message. If the endpoint includes **transportQoS** in an ARQ message, the capabilities specified apply only to that particular call.

The **transportQoS** field is an optional parameter in an ARQ message. It indicates whether the endpoint intends to participate in transport-QoS exchange and implicitly also indicates whether such functionality is required. The parameter can indicate one of the following:

- Endpoint Controlled: This option implies that the endpoint will control the transport QoS exchange.
- Gatekeeper Controlled: In this option, the endpoint expresses that the gatekeeper control the transport QoS exchange on behalf of the endpoint.
- No Control: This option implies that no QoS exchange is necessary. This option indicates to the gatekeeper that QoS exchange is not necessary for the call.

While the above provides the gatekeeper the endpoint's selection with QoS control, it does provide much indication about the capability of the endpoint itself. Therefore, a new parameter for **transportQoS** is recommended:

- **QoSCapable** (Boolean): The endpoint can specify whether or not it is QoS capable in the **transportQoS** parameter. This will enable the gatekeeper decide whether to request the endpoint to control QoS in the event that the endpoint had specified gatekeeper controlled or no control.

[Author: Strength tag may also be need to be included.]

7.1.2 Gatekeeper Selection of QoS capabilities

The gatekeeper decides whether to accept or reject the QoS capabilities received in the ARQ message based on the received information and the information it has about the state of the network, any defaults configured etc. The gatekeeper indicates its decision by inserting a **transportQoS** parameter in the ACF ARJ, RCF, or in an RRJ message. The gatekeeper should include the endpoint’s QoS capabilities in all the subsequent LCF messages that are sent.

The options are as follows:

- **Endpoint Controlled:** In an ACF message, this confirms that the endpoint’s control of Transport-QoS. In an ARJ, it rejects the indication in an ARQ and suggests that the endpoint should control transport-QoS.
- **Gatekeeper Controlled:** In an ACF message, this confirms that the gatekeeper’s control of Transport-QoS. In an ARJ, it rejects the indication in the ARQ and suggests that the gatekeeper should control transport-QoS.
- **No-Control:** If included in an ACF messages, then it is an indication that the no-control of transport-QoS is required. However, if included in an ARJ messages, then it implies that no control of transport-QoS is required and therefore rejects the suggestion made in the ARQ.

If relayed in the RCF message, the decision applies to all calls made by the endpoint, unless the gatekeeper later supplies a **transportQoS** field in an ACF message. If relayed in the ACF/ARJ message, the decision applies only to one particular call. Table 4 lists the various options and gatekeeper responses. The endpoint shall accept the gatekeeper's decision in order to place a call.

Endpoint Choices	Gatekeeper Choices		
	Endpoint-Controlled	Gatekeeper-Controlled	No-Control
Endpoint-Controlled	ACF/RCF	ARJ/RRJ (gatekeeper-control)	ARJ/RRJ (no-control)
Gatekeeper-Controlled	ARJ/RRJ (endpoint-control)	ACF/RCF	ARJ/RRJ (no-control)
No-Control	ARJ/RRJ (endpoint-control)	ARJ/RRJ (gatekeeper-control)	ACF/RCF

Figure 4: Gatekeeper's responses in transportQoS

[Author: The Gatekeeper controlled services structure and discussion, present in the previous version, has been removed temporarily. The intention is to add them back after discussion.]

7.1.3 Endpoint bandwidth requests

*[Author: The **BandWidth** detail structure and discussion, present in the previous version, has been removed temporarily. This can easily be added back if the general consensus requires it.]*

In addition to **transportQoS**, an endpoint should also calculate and report the bandwidth it currently intends to use for all channels of the call. The **bandWidth** parameter is used for such purpose. This bandwidth request should be reported in the ARQ message. The bandwidth parameters reported in the ARQ message is independent of the decision by the endpoint or by the gatekeeper to use any network level resource reservation schemes (e.g., RSVP in the case of the IP network, use of QoS Classes at the time of call setup in the case of ATM network) or not.

In addition, if bandwidth requirements change during the course of the call, an endpoint should report changes in bandwidth requirements to the Gatekeeper using the BRQ.

7.2 Call Setup Procedures

Currently, the alerting of the callee happens during the call-setup phase whereas the media stream setup occurs much later. This causes the QoS establishment to occur after the alerting phase which may lead to undesirable user scenarios depending on the **qosType** chosen. With "required" **qosType**, QoS establishment must occur before alerting. There are 2 ways in which QOS establishment can occur before the alerting of the callee. They are:

- Fast Start: Fast start procedures allow for the QoS negotiation during call-setup which in-turn allows for QoS establishment before the alerting phase.
- Inclusion of the H.245 address in the Setup message
- H.245 Tunnelling

If the receiving H.323 entity requires a "required" **qosType** and it receives the Setup messages without any of the above then the call establishment is failed

7.2.1 Fast-Start Procedures

Fast-start procedures may be used by the H.323 entities to enable QoS establishment before pre-ring. In this procedure, a sequence of OpenLogicalStructures is included in the Setup message. To allow for QoS negotiation, these procedures contain the QoS parameters as well. The presence of the GBD indicates to the receiving H.323 entity that QoS procedures are required. This enables the withholding of the alerting until the QoS procedures are complete.

7.2.2 H.245 Address in the Setup message

In this mechanism, the H.323 entity adds the H.245 address in the Setup message. Once the other party receives the H.245 address, it can initiate the H.245 exchange which allows for QoS negotiation. Until the QoS negotiation is complete, alerting is withheld. A Call Proceeding message is sent to prevent timeouts.

7.2.3 H.245 Tunnelling

H.245 tunnelling is another mechanism by which the H.245 information necessary for the QoS procedures can be exchanged during the call-setup process. This allows an endpoint to initiate the QoS procedures and ensures the requested QoS is available before the alerting process.

7.3 Bearer/Media Stream Setup Procedures

The above section dealt with how the H.245 exchange can be made possible during the call-setup phase. This section elaborates the QoS handling within the H.245 exchange.

7.3.1 H.245 Capability Exchange Phase

During the H.245 capability exchange, each endpoint can indicate it's transmit and receive QoS capabilities to the other endpoint. This exchange is done using the GBD parameter. The H.245 capability exchange is not stream specific but represents a cumulative required for all streams. Hence, there is no value in inserting information beyond the **qosType** in this exchange. An absence of this parameter indicates no QoS is necessary to the other H.323 entity. After capability exchange, the open logical channel phase is used to appropriate resources that are specific to a given medium.

7.3.2 Logical Channel Signalling

In this phase, the opening of the H.245 logical channel is where the main QoS exchanges happen and reserving resources are done. Reservations (guaranteed or controlled) are performed only if both H.323 endpoints indicate that they are H.323 QoS is enabled during capability exchange.

The **qosType** indicates whether the call can proceed even if the QoS request failed. If even a single call leg is "required" then the following rules apply:

- A stream is said to have "required" QoS even if one of the call-legs has a "required" policy. At each H.323 entity, the **qosType** is combined with the **qosType** from the incoming message to derive the **derivedQoSType**. The **derivedQoSType** is the one that apply and is also used in the QoSDescriptor that is forwarded onwards. If a "required" **qosType** is combined with a "desired" **qosType**, then the resulting **derivedQoSType** is "required". A "required" **qosType** in any call leg will cause a stream to be failed if QoS is not secured for any leg of the call.
- Any H.323 entity that receives a QoS request rejection should initiate the call teardown in the case of the "required" QoS.
- The callee must not alert the user until the confirmation to the QoS request is received. This is done to avoid a situation where the user is alerted and the call is failed subsequently.
- All the rules above apply to a single logical channel (stream). The H.323 entities must contain policies that dictate what kind of coordination is required between the streams. For example: A QoS failure in a video stream may not need the failure of the call. The call may be continued as an audio call or a call with no video just on one direction. Whereas, a failure in the audio stream may cause a complete call failure.

7.4 Authorization Procedures

The authorization process depends upon each domain. The authorization processes occur in the H.323 system as well as in the transport system. In the application plane, the **QoSM**, along with **QoSPE**, authorize the call and ensure that the end-user/call is allowed to ask for the level of QoS requested, i.e. the **servicePriority**, the **qosType**, the **qosValues**, etc are all within the limits allowed for the particular end-user/call.

In the transport system, the network device must be assured that the resources requested are indeed allowed by the H.323 system. The authorization of the network usage can be done in a couple of ways. One is a push-down model, where the **QoSM** pushes down the authorization to the relevant network entities. In the other model, the pull model, the network entity on being requested for resources, queries the application plane regarding the validity of the request.

The problem with the push-down model is that the application plane must learn the identity of the relevant network device which is going to service the bearer stream. This is a difficult task in heterogeneous networks. The advantage of this model is however that authorization can be done in parallel and has no delay implications.

The problem of the pull model is that it may add to delay since the polling has to be done during resource setup. Also, since all transport-QoS architecture does not support out-of-band QoS signalling, this model cannot be deployed in all domains.

The architecture and process laid out in this document shall allow either model to be deployed. It is also foreseen that multiple models may be deployed in a single call.

8 H.225.0 Annex G – Communication between Administrative Domains

[Editor's comment: Need to consider whether this text should be in Annex G with other Annex G procedures instead of in H.323.]

The **transportQOS** field is used in the Descriptor message for inter-domain communications. Like intra-domain communications, the **transportQOS** field will also indicate whether the endpoint or the gatekeeper will have the capability to reserve transport resources in accordance with guaranteed and controlled services criteria. If none of these capabilities exists, it will indicate unspecified services (i.e., best effort) criteria. The capabilities of the guaranteed, controlled, and unspecified services criteria will mean that resources can be reserved in accordance with the QoS parameters described in Appendix I.

9 References

- AT&T, VocalTec, Alcatel, Motorola and 3Com, "Call Flows Supporting H.323 QoS," APC-1668. ITU-T SG16, Q.11-15/16 Joint Rapporteurs' Meeting, Red Bank, New Jersey, USA, October 18-22, 1999.
- AT&T, VocalTec, Alcatel, Motorola and 3Com, "Extensions of H.225.0 and H.225.0 Annex G Signaling Messages and their ASN.1 Syntax to Support H.323 QoS," APC-1669. ITU-T SG16, Q.11-15/16 Joint Rapporteurs' Meeting, Red Bank, New Jersey, USA, October 18-22, 1999.
- AT&T, VocalTec, Alcatel, Motorola and 3Com, "Framework for Mapping of H.323 QoS over Packet-based Networks," APC-1670. ITU-T SG16, Q.11-15/16 Joint Rapporteurs' Meeting, Red Bank, New Jersey, USA, October 18-22, 1999.
- Nokia, "On QoS for H.323 Systems," APC-1705. ITU-T SG16, Q.11-15/16 Joint Rapporteurs' Meeting, Red Bank, New Jersey, USA, October 18-22, 1999.
- AT&T, "H.323 Differentiated Services and Their Protocol Architectures," APC-1492. ITU-T SG16, Q.11-15/16 Joint Rapporteurs' Meeting, Monterey, CA, USA, February 15-19, 1999.

- AT&T, “An enhancement mechanism for differentiation of H.323 Quality-of-Services (QoS),” D.222, ITU-T SG16 Meeting, Santiago, Chile, May 17-28, 1999.
- AT&T, VocalTec Communications, DataBeam/IBM, and Alcatel, “Functional Requirements and Service Classes to Support QoS in H.323,” APC-1591, Study Group 16 Q.11-15 Joint Rapporteurs’ Meeting, Berlin, Germany, 2-6 August, 1999.
- “Liaison to TIPHON, IETF, and ATM FORUM on Functional Requirements and Service Classes to Support QoS in H.323, for information and comment,” TD-17, Study Group 16 Q.11-15 Joint Rapporteurs’ Meeting, Berlin, Germany, 2-6 August, 1999.

Appendix I

10 Mapping of H.323 QoS classes over specific transport layers

10.1 Objectives of QoS mapping

This section defines the mapping of application level H.323 QoS classes to specific network layer QoS mechanisms defined by packet- and cell- based networks such as IP and ATM.

The varied network conditions in a packet network could result in congestion, unacceptable packet/cell delays or too much packet/cell delay variation. In order to mitigate these problems for the traffic flows, various network services have been standardized by the IETF. These include Integrated Services, Differentiated Services, Resource Reservation Protocol (RSVP), MPLS etc. They are similar issues in ATM environment as well.

An end user may request any H.323 QoS class taking into considering the cost-performance trade-offs. These requirements are needed end-to-end. As an end-to-end path may support more than one network layer service, the H.323 QoS requirements will need to be translated from one network layer service into another. This network layer translation could lead to inconsistent QoS services end-to-end. Therefore, a mapping function between the H.323 QoS classes and the network layer services are required to deliver required H.323 QoS services end-to-end. The example below will clarify further:

エラー! 参照元が見つかりません。 shows that the three administrative domains with different network layer QoS services that offer H.323 services The H.323 protocol is common to all the three domains which enables an application such as VOIP across these domains.

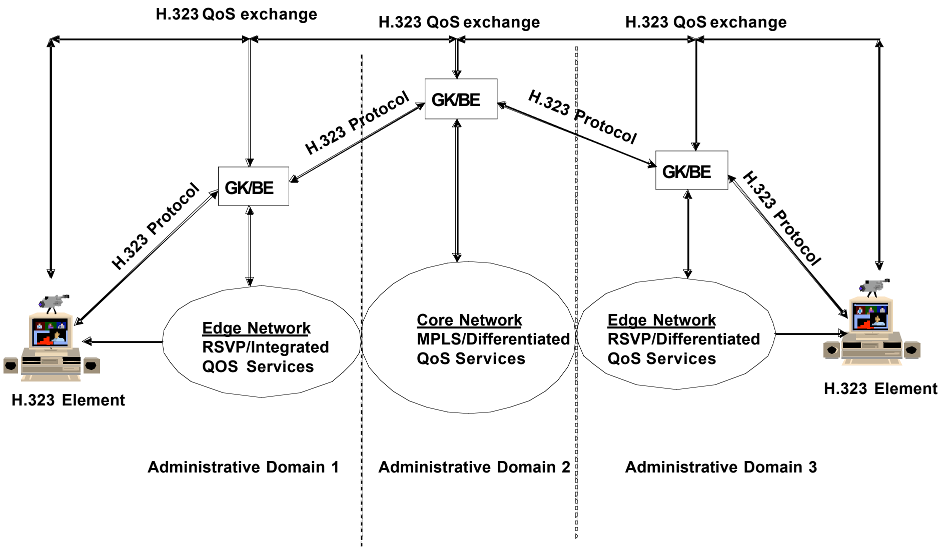

Figure 5: Domains with Different Network Layer Services

The H.323 terminals and gatekeepers/border elements in all the 3 domains communicate using the common H.323 protocol. The call is negotiated with a required H.323 QoS class that are understood by all H.323 entities of all domains. If the H.323 QoS class has to be translated from one domain to another in the network layer, then the resultant end-to-end QoS services may not match the application's request. Therefore, a mapping function between the application and the network layer services is required. As the application layer QoS classes are mapped evenly to all the different network layer services in the path, a consistent QoS services becomes possible end-to-end.

[Author: The actual mapping is not provided yet. The exchange of the QoS class is to be discussed after which this section will be completed.

10.2 IP Network QoS

The IETF has introduced Integrated services (IntServ/RSVP), Differentiated services (DiffServ), Resource ReSerVation Protocol (RSVP), and MultiProtocol Label Switching (MPLS) to meet the demand for QoS in IP networks.

IntServ is a QoS architecture which allows for path-based admission control. RSVP is the signalling protocol used in this architecture. In IntServ architecture, classification, policing and scheduling occurs in conjunction with the admission control process.

DiffServ is class-based QoS architecture which supports in-band signalling. The signalling occurs via a value defined in the type of service (TOS) byte (also called differentiated services (DS) field) of the IP header. This value is referred to as the Differentiated Services code point (DSCP). The packet forwarding treatment given to a packet is based on the DSCP value and is known as the per-hop behaviour (PHB).

RSVP is a QoS signalling protocol that allows applications to reserve network resources. RSVP is a flow-based QoS mechanism which can be used for individual flows as well as flow aggregates. RSVP is traditionally used in IntServ architecture but may be used with DiffServ architecture as well. RSVP with DiffServ does not have the scalability concerns that went with IntServ RSVP. RSVP aggregation is a RSVP mechanism in which reservations are aggregated and treated as a single unit. This aggregation solution is a very scalable solution for networks with heavy load and is therefore suitable for large networks.

MPLS uses a fixed length label ahead of the IP address. This label functions as an index for the selected route. The class-of service (COS) bits within the short fixed-size label provides the indication of the QoS for the packet.

10.3 Mapping H.323 QoS to IntServ/RSVP

RSVP is a QoS signalling protocol that enables applications to request reservation of network resources. These requests dictate the level of resources (e.g., bandwidth, buffer space) that must be reserved along with the transmission scheduling behaviour. The transmission scheduling behaviour must be installed in the network layer devices (e.g., routers) to provide the desired end-to-end QoS commitment for the data flow. The QoS can be provided on a per-flow basis according to requests from the end application.

RSVP may be deployed in two ways. One is a pure IntServ approach where RSVP acts not only on the control plane providing admission control but is also used on the data plane providing the policing, queuing and scheduling of the flow. This was the original model of RSVP. However, as the per-flow state information with RSVP increases proportionally with the number of flows, it causes storage and processing overhead on the routers. To address this issue, the control plane and the data plane actions in RSVP were separated in the IntServ/DiffServ approach, which is defined in RFC 2998. RSVP acts on the control plane and allows DiffServ processing in the data plane. This has helped alleviate some of the scaling concerns.

For higher scalability, RSVP aggregation may also be deployed. In this mechanism, the RSVP reservations are aggregated together and the admission control is performed on the aggregated reservation. For the quantitative applications, RSVP offers a “guaranteed” and a “controlled” service to the network.

The guaranteed service is for real-time applications that are unable to handle delay – it attempts to deliver a practicable, constant stream of network capacity that is as close as possible to the end-to-end network delay.

The controlled-load service is a better than best-effort service; it attempts to deliver end-to-end network capacity as close as possible to the condition of an unloaded network, but still best effort. Controlled-load contracts agree that a flow will be handled within a certain range, but variance is anticipated. It is not expected to accept or use specific values for control parameters that include information about delay or loss.

This mechanism involves the exchange of RSVP messages between the 2 endpoints. These messages are forwarded along the same route as the bearer stream and so admission control is done on the same devices/interfaces which service the stream. It is not necessary to enable RSVP on every hop. RSVP messages are merely forwarded along in devices where RSVP is not enabled. One possible deployment may be that RSVP is enabled at the edge of network and WAN links whereas the other links may continue to have DiffServ QoS. The RSVP messages carry sufficient information to identify the stream such as source and destination IP address and port. They also contain the traffic profile and parameters that describe the required QoS. Each RSVP-enabled device either grants or denies the request based on the available capacity in the interface and the policies configured. For example: A request may be denied even when sufficient capacity exists if policy control fails. Policies may be based on user and application information or some additional credential.

In RSVP, traffic can be characterized by bucket rate/service rate/bandwidth (bytes per second), peak rate of flow (bytes per second), bucket depth/maximum datagram size/maximum burst size (bytes) and other parameters. The required QoS is described by token slack term/delay (milliseconds) and variation in delay.

Once the request is granted, the required resources in the interface is granted and kept aside for the flow. The bearer stream may be processed via the IntServ method or the DiffServ method as described above to provide QoS.

10.3.1.1.1 Mapping H.323 QoS to DiffServ

The IETF’s DiffServ is an inband-signalling mechanism that is relatively simple to implement. From the development point of view, the mechanism calls for the appropriate DSCP value marking in the IP header of the packet. It is a very scalable solution and can be deployed in network clouds where there is very high traffic load. The disadvantage of this mechanism is that there is no explicit admission control and no feedback to the user when adequate QoS cannot be granted.

DiffServ requires that some capacity in the network be set aside for particular classes of traffic. In this mechanism, a set of primitives are applied to the traffic. They are: Classification, Policing, Shaping and Marking. Classification is done based on the DSCP values contained in the IP header of the packet. The DSCP value is a 6 bit value and therefore can range from 0 to 63. Some of the values within this range are defined by IETF standards and their associated per-hop behaviour is outlined as well. There are some values within that range that are left for experimental purposes. DiffServ policing primitives will police the traffic based on the given profile. If traffic within a class exceeds the given profile, then there are either dropped or shaped. Shaping primitives causes the traffic to be delayed and forwarded rather than dropped when the profile max has been reached. Shaping also helps in smoothing the flow of packets within a class. Finally the marking primitives will mark or remark the packets based on the given DSCP value.

Using these above mechanisms and others QoS tools, the DiffServ method can provide a variety of services such as premium service for applications requiring low-delay and low jitter service, assured service for application requiring better reliability that best-effort service and others.

The increased scalability in the DiffServ method is due to the following:

- Classification using just the DiffServ values
- Limited state information as state is maintained per class and not per flow
- All primitives are not required in every hop. One example could be that all primitives are employed at the edge while just the classification primitive is employed at the core routers.

The disadvantage of this solution is that there is no protection of traffic within a class. For example: If too many packets arrive at the router for admission within the same class, all the excess packets will most likely be dropped. This causes quality degradation across all traffic flows. Instead, with an explicit admission control, only a smaller set of flows would have been affected while the other flows would be provided with a guaranteed quality. The other major disadvantage of the diffserv approach is that there is no notification back to the application on such losses.

Appendix II

Modifications to H.225.0 v5

[Editor: This section will be worked on when consensus is reached on the above sections.]