Telecommunication Standardization Sector

Original: English

(TSS)

AVC-822
also AVC-757 revision 4

Experts Group for Video Coding and Systems in ATM and Other Network Environments

October 23-28, 1995

STUDY GROUP 15
CONTRIBUTION

Source: D. Skran, Editor, H.22Z.
email: dls@mtgbcs.att.com
voice: +1 (908) 957-5988
fax: +1 (908) 957-5627

Title: H.22Z, Media Stream Synchronization and Time Base recovery on Non-Guaranteed Quality of Service LANs (Suggested alternative is "Multiplexing Protocol for Visual Telephone Systems on Non-Guaranteed Quality of Service LANs")

Date: October 20~~August 19~~, 1995

Summary: This recommendation describes a method for combining ~~how~~ audio, video, data, and control information on a non-guaranteed quality of service LAN ~~can be managed~~ to provide conversational services in H.323 equipment. Included are such topics as audio coding, video coding, control and signaling PDUs, and methods for providing improved quality of service in this environment.

Notes on reading: ***{Editors notes are generally in braces}*** while underlined text is new. ~~The deleted text has strikethrough font.~~ Please ignore all references to other sections both inside and outside this document; these will be updated in the final version.

This version (rev 4~~3~~) reflects changes from the August 19~~July 31~~ version(rev 3~~2~~ after Haninge).. It is the fourtHh~~third~~ version to attempt to make use of RTP/RTCP more fully. This version includes comments from Terry Lyons/AT&T (9/11), Vineet Kumar/Intel (10/12), Mark Reid/PictureTel(9/8), and Ami Amir/RadVision(9/12). No attempt has been made to fully integrate all the proposed PDUs into this version.

Issues:

1)Add maximum packet rate to cap exchange?
2)Use Q.931 and if so how?
3)MB coding and hardware coders

INTERNATIONAL TELECOMMUNICATION UNION

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

DRAFT H.22Z

(October 20, 1995)

LINE TRANSMISSION OF NON-TELEPHONE SIGNALS

Media Stream Synchronization and Time Base recovery on Non-Guaranteed Quality of Service LANs

DRAFT ITU-T Recommendation H.22Z

FOREWORD

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

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

The approval of Recommendations by the Members of the ITU-T is covered by the procedure laid down in WTSC Resolution No. 1 (Helsinki, March 1-12, 1993)

ITU-T Recommendation H.22Z was prepared by the ITU-T Study Group 15 (1993-1996) and was approved under the WTSC Resolution No. 1 procedure on the xxth of xxxx 199x.

NOTE

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

SUMMARY

This Recommendation covers the technical requirements for narrow-band visual telephone services defined in H.200/AV.120-Series Recommendations, in those situations where the transmission path includes one or more Local Area Networks (LAN), each of which is configured and managed to provide a non-guaranteed Quality of Service (QoS) which is not equivalent to that of N-ISDN such that additional protection or recovery mechanisms beyond those mandated by Rec. H.320 need be provided in the terminals. It is noted that Recommendation H.322 addresses the use of some other LANs which are able to provide the underlying performance not assumed by the H.323/H.22Z Recommendations.

This recommendation describes how audio, video, data, and control information on a non-guaranteed quality of service LAN can be managed to provide conversational services in H.323 equipment

Media Stream Synchronization and Time Base recovery on Non-Guaranteed Quality of Service LANs

(Place, 199x)

The ITU,

considering
the widespread adoption of and the increasing use of the H.320 Recommendation for videophony and videoconferencing services over networks conforming to the N-ISDN characteristics specified in the I series Recommendations,

appreciating
the desirability and benefits of enabling the above services to be carried, wholly or in part, over Local Area Networks while also maintaining the capability of interworking with H.320 terminals

and noting
the characteristics and performances of the many types of Local Area Network which are of potential interest

recommends
that systems and equipment meeting the requirements of the H.322 or H.323 Recommendations are utilized to provide these facilities.

1. Scope

This recommendation describes the means by which audio, video, data, and control are associated, coded, and packetized for ~~synchronized~~ transport~~and passed~~ between H.323 terminals on a non-guaranteed quality of service LAN, or between H.323 terminals and an H.343 gateway, which in turn may be connected to H.320, H.324, or H.310/H.321 terminals on N-ISDN, GSTNGSTN~~PSTN~~, or B-ISDN respectively~~-via a LAN/WAN gateway~~. This gateway, terminal descriptions, and procedures are described in H.323 while H.22Z covers protocols and message formats. Communication via an H.323 gateway to an H.322 gateway for guaranteed quality of service (QOS) LANs and thus to H.322 endpoints is also possible.

H.22Z is intended to operate over a variety of different LANs, including IEEE 802.3, Token Ring, etc.. In this context, "LAN" includes the transport layer, such as TCP/IP/UDP, SPX/IPX, etc. Thus, H.22Z is defined as being above the transport layer. It is expected that a convergence sublayer would exist between H.22Z and the underlying real LAN. Many characteristics of this sub-layer are left to the manufacturer, but some elements require standardization for interoperability between H.323 terminals on the same LAN type. Specific profiles for particular LAN protocol suites are included in Annex A of this recommendation. Thus, the scope of H.22Z communication is between H.323 terminals and H.323 gateways on the same LAN, using the same convergence sub-layer. This LAN may be a single segment or ring, or it logically could be the an enterprise data network comprising multiple LANs bridged or routed to create on continuous network~~entire Internet~~. It should be emphasized that operation of H.323 terminals over the entire Internet, or even several connected LANs may result in poor performance. The possible means by which quality of service might be assured on this LAN, or on the Internet in general is beyond the scope of this recommendation. However, H.22Z provides a means for the user of H.323 equipment to determine that quality problems are the result of LAN congestion, as well as procedures for corrective actions. It is also noted that the use of multiple H.323 gateways connected over the public ISDN network is a straightforward method for increasing quality of service.

H.323/H.22Z are intended to extend H.320/H.221 conferences/connections onto the non-guaranteed QOS LAN environment. As such the primary conference model is one with size in the range of a few participants to a few thousand, as opposed to large-scale broadcast operations, with~~with size less than 1000 parties,~~ strong admission control, and tight conference control. This is in contrast to various IETF (Intern~~N~~et Engineering Task Force) protocols that focus on very large conferences with weak admission and conference control.

H.22Z makes use of the IETFs RTP/RTCP(Real Time Protocol/Real Time Control Protocol) for media steam packetization and synchronization for all underlying LANs. Please note that the usage of RTP/RTCP as specified in H.22Z is not tied in any way to the usage of TCP/IP/UDP. H.22Z assumes a call model where initial signaling on a non-RTP LAN port is used for call establishment and capability negotiation (see H.323 and H.245), followed by the establishment of one or more RTP/RTCP connections. H.22Z contains details on the usage of RTP/RTCP~~, and also specifies any extensions to RTP/RTCP that are required for interworking with H.320~~.

Since H.22Z supports only single connection operation in the H.221 sense (i.e. any "channel" structure provided by the LAN is analogous to the "channels" of H.221, such as audio, video, LSD, etc. Thus, on the LAN side, for example, there are only single "channel"~~128 kbps~~ calls with a maximum rate limited to 128 kbps, not 2 * 64 kbps fixed rate calls)~~; this rate matching function requires a little more thought~~. Another example has single "channel" LAN calls with a maximum rate limited to 384 kbps interoworking with 6 * 64 kbps on the WAN side[1]. -The primary rationale is to put complexity in the gateway rather than the terminal and to avoid extending onto the LAN features of H.320 that are tightly tied to ISDN unless this is absolutely necessary.

{Editors Note: There are two possible ways to deal with the rate matching problem:

1)extend the transfer rate back to the H.323 terminal. For example, this implies signaling a 2*64 call differently than a 128 kbps call, since they differ in video rate by 1600 bits/sec, and requiring the H.323 terminal to follow whatever the H.320 rate is.

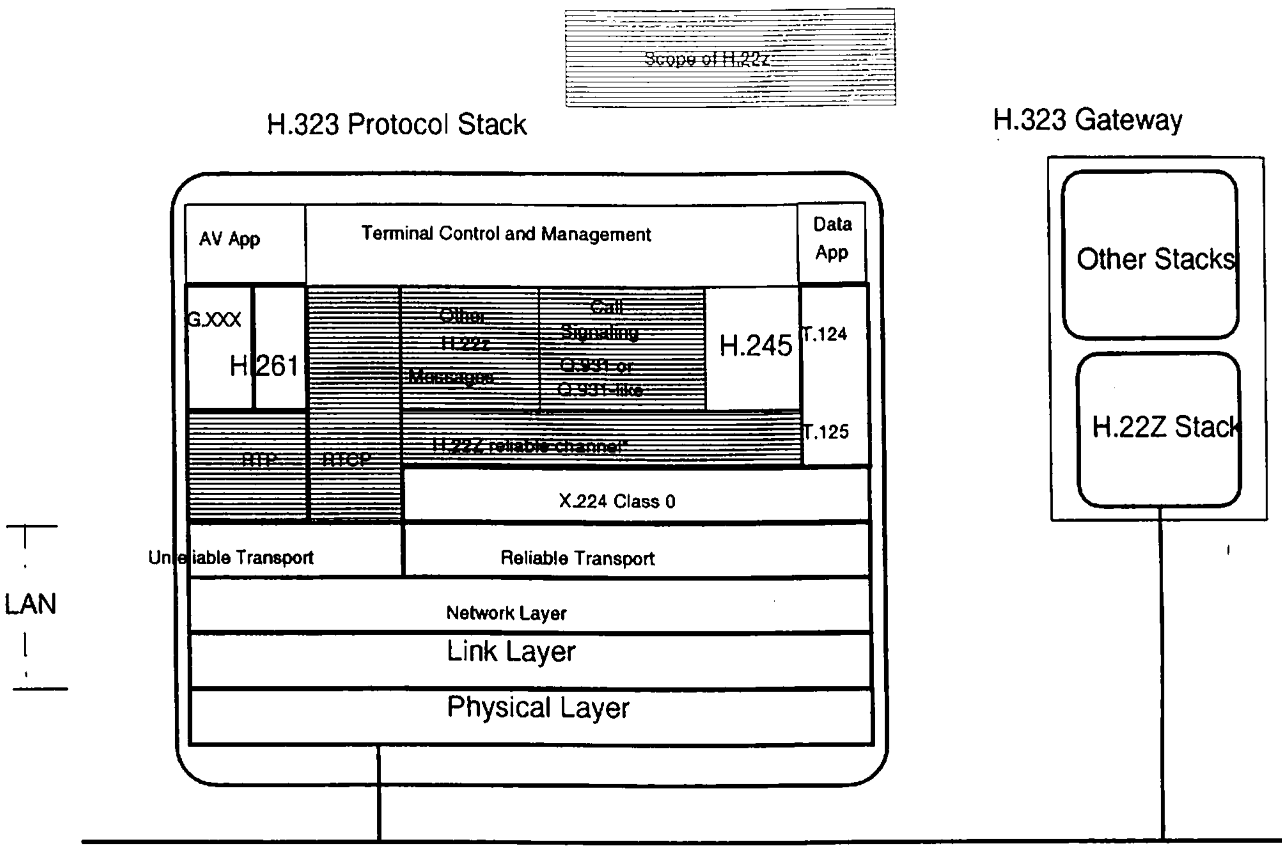
2)hide the differences in transfer rate on the H.320 side from the H.323 terminal. Thus, the gateway might perform a rate reduction or bit stuffing function.

Initiallly, I thought that having the gateway act as an H.244 CAU might be helpful, but at this moment, item (1) above seems simpler. Comments?

As best I recall, the consensus of comments was for option (1). This will require a means to signal the desired rate either in H.245 or in the H.22Z control channel. }

H.22Z is designed so that, with an H.323 gateway, interoperability~~interpretability~~ with H.320(1990), H.320(1993), and H.320(1996) terminals is possible. However, some features of H.22Z may be directed toward allowing enhanced operations with future versions of H.320. It is also possible that the quality of service on the H.320 side may vary based on the features and capabilities of the H.323 gateway.

[1]Note that video plus data rates on the LAN side must match the video plus data rates in the WAN side H.320 multiplex; the audio and control rates are not required to match. Also note that the LAN rate may under-run the WAN rate for either/both video or/and data, but it cannot exceed the maximum amount that fits into the WAN side multiplex.



* Unreliable channel to Gatekeeper for inital H.22Z signaling channel is not shown since it is not to the distant terminal

Figure 1/H.22Z
Scope of H.22Z

The general approach of H.22Z is to provide a means of synchronizing packets that makes use of the underlying LAN facilities. H.22Z is not a "multiplex" and does not attempt to pack all media and control into a single stream, which is then packetized. The framing mechanisms of H.221 are not utilized for the following reasons:

- Not using H.221 allows each media to receive different error treatment as appropriate.
- H.221 is relatively sensitive to the loss of random groups of bits; packetization allows greater robustness in the LAN environment.
- H.245 and Q.931 (or Q.931-like call related messages) can be sent over a reliable link provided by the LAN.

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; all 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.

{Editors Notes: References are included to RTP/RTCP and related documents to avoid duplication of text in H.22Z. Our current focus is on establishing a structure for the usage of RTP/RTCP. ***~~The issue of how to include the IETF documents is deferred for the time being. Options appear to be:~~***

~~a)Include by reference - it is unclear that the ITU-T can include an IETF RFC as a normative document~~

~~b)Include by reference but in a non-normative annex. This has the potential to lead to interoperability problems~~

~~c)Include the text of RTP in H.22Z, making RTP an ITU-T document~~

~~d)Create an H.RTP, H.RTCP, etc. as needed, and refer to these documents to avoid making H.22Z too large; note that the RTP related documents total 94 pages!.~~

~~e)Decide to not use RTP/RTCP after all.~~

~~Both (c) and (d) have the potential to lead to divergence between the ITU-T and IETF versions.~~

<u>The use of RTP/RTCP will be hopefully finally settled at the October meeting. At this time, it appears that RTP has many supporters and one major opponent. Thus, I have not deviated from the RTP course since it seems to be supported by the great majority of views expressed so far.</u>

<u>Based on information I have received, it appears that referencing RTP/RTCP is allowed by the ITU. Thus, I propose that at the November meeting we send a liaison (possibly drafted at the October meeting) to the IETF, indicating that we intend to use RTP/RTCP and associated documents, and that we desire that by May 1996 we have a fixed version number to refer to. Unfortunately, I am informed that this may be difficult for the IETF to agree to, as their "standards" process may well continue through the end of 1996. It is unclear to me how to deal with this situation; however we could refer to an RFC as long as it would be absolutely fixed as a reference point by May of 1996. Alternatively, we could delay H.22Z, possibly adopting H.323 in May. Unfortunately, this would delay the adoption of H.22Z until late 1997. The above possibilities still exist as well.}</u>

1. CCITT Recommendation G.711, *Pulse Code Modulation of 3kHz Audio Channels*, November 1988

2. CCITT Recommendation G.722, *7kHz Audio Coding within a 64 kbit/s Channel*, November 1988}

3. CCITT Recommendation G.728, *Coding of Speech at 16 kbit/s Using Low-delay Code Excited Linear Prediction (LD-CELP)*, May 1992

4. ITU-T Recommendation H.221(1993)[2], *Frame Structure for a 64 to 1920 kbit/s channel in audiovisual teleservices.*

[2]Previously CCITT Recommendation

5. ITU-T Recommendation H.230(1993)[3], *Frame Synchronous Control and Indication Signals for Audiovisual Systems,* December 1990

6. ITU-T Recommendation H.233(1993) *Confidentiality System for Audiovisual Services,* March 1993

7. ITU-T Recommendation H.242(1993)[4], *System for Establishing Communication Between Audiovisual Terminals Using Digital Channels up to 2 Mbit/s.*

8. ITU-T Recommendation H.243(1993), *Procedures for Establishing Communication Between Three or More Audiovisual Terminals using Channels up to 2 Mbit/sec.*

9. ITY-T Recommendation H.320(1993)[5], *Narrowband Visual Telephone Systems and Terminal Equipment.*

10. ITU-T Recommendation T.122(1993), *Multipoint Communication Service for Audiographics and Audiovisual Conferencing Service Definition*

11. ITU-T Recommendation T.123(1993), *Protocol Stacks for Audiovisual and Audiographic Teleconference Applications.*

12. ITU-T Recommendation T.125(1994), *Multipoint Communication Service Protocol Specification.*

13. H.324.

14. H.321

15. H.322

16. H.310

17. RTP: A Transport Protocol for Real-Time Applications, IETF, Version 2, March 21, 1995

18. RTP Profile for Audio and Video Conferences with Minimal Control, IETF, July 7, 1995

19. RTP Payload Format for H.261 Video Streams, IETF June 9, 1995~~05~~

3. Definitions

Internet: Any system for the connection of LANs at the network layer. The operation of such systems is beyond the scope of H.22Z.

LAN: In the context of this document, a LAN(Local Area Network) is a mechanism for the switching of packet data over a limited area, including the physical, link, and transport layers.

[3]Previously CCITT Recommendation
[4]Previously CCITT Recommendation
[5]Previously CCITT Recommendation

Thus, the "LAN" is beyond the scope of H.22Z for the most part, and H.22Z is intended for use with a variety of "LANs."

LAN Port: A destination on the LAN with a known address. May be multicast or unicast. Not to be confused with an MCU port. *{Editor's note:~~- Is there a better term?~~ Based on various inputs, I suggest the following:*

LAN Channel: The combination of a LAN Network Address and a Port number. The channel may be multicast or unicast. In H.22Z each media uses a different LAN channel. Not to be confused with an ISDN channel.

LAN Network Address: The addressing entity by which a terminal is referred to externally. An example is the IP address.

LAN Port Number: a value that when associated with a LAN Network Address makes up a LAN channel. A Port Number may be specified in various ways for differnt LANs, but it is that entity which allows sub-addressing within a LAN Network Address. The LAN Port number may provide guaranteeed delivery or non-guaranteeded delivery of data. Not to be confused with an MCU port.

Comments? I have not spent much time updating the document so far pending further input.
}

Multipoint Control Unit (MCU): a multi-port device, by means of which two or more audiovisual terminals may intercommunicate in a conference call; See recommendations H.231 and H.243 for details. The MCU is used to connect terminals over a wide area, and especially to connect two or more terminals entering the WAN via an H.323 gateway into a WAN conference.

See H.323 for definitions of domain, super-domain, and sub-domain.

4. Conventions

In this document, "shall" refers to a mandatory requirement, while "should" refers to a recommended but optional feature or procedure.

5. Abbreviations

BAS	Bit rate Allocation Signal
CIF	Common Intermediate Format
ECS	Encryption Control Signal
FAS	Frame Alignment Signal
FAW	Frame Alignment Word
~~**GLI**~~	~~General Lan Interface~~
GOB	Group of Blocks
H-MLP	High speed Multi-Layer Protocol
HSD	High Speed Data

IETF	Internet Engineering Task Force
IP	Add
LAN	Local Area Network
LD-CELP	Low Delay - Code Excited Linear Prediction
LSB	Least Significant Bit
LSD	Low Speed Data
MBE	Multi-Byte Extension
MCC	Multipoint Command Conference
MCN	Multipoint Command Negating
MCS	Multipoint Command Data Symmetrical Data Transmission
MF	MultiFrame
MLP	Multi-Layer Protocol
MPI	Minimum Picture Interval
MSB	Most Significant Bit
NS	Non-Standard
PCM	Pulse Code Modulation
QCIF	Quarter Common Intermediate Format
QOS	Quality of Service
RTP	Real Time Protocol
RTCP	Real Time Control Protocol
SBE	Single Byte Extension
SC	Service Channel
SCM	Selected Communications Mode
SMF	Sub-MultiFrame
SP	Still Picture
TCP	Add
VCF	Video Command "Freeze Picture Request"
VCU	Video Command "Fast Update Request"
UDP	Add

{Editor's Note: It is proposed that the GLI be removed. I have received some input that it is not useful. I am also tending toward the idea that it does not add any value that is not already in Annex A. Comments?

6. Multiplex~~Synchronization~~ Mechanism

6.1. General Approach

The general approach is to send an initial connect request on a non-guaranteed channel. As a result of this request, a channel with guaranteed delivery is established for both additional call management~~setup~~ messages ~~(Q.931)~~ and control(H.245) ~~using an underlying guaranteed delivery means of transport~~ . The impact of varying delay on these types of PDUs is not as strong as the impact on audio and video. These messages are contained in a control PDU structure described in section XXX below. They are used to first establish the connection between the gateway (and optionally the gatekeeper) and the terminal ~~(Q.931)~~, and then to exchange capabilities (H.245). In circumstances in which the gateway/gatekeeper are not required, this method is also used for point-to-point LAN call setup.

Once this guaranteed delivery control LAN port has been established, additional LAN ports for audio, video, and data may be established based on the outcome of the capability exchange. Also, the nature of the LAN side multi-media conference (centralized vs distributed/multicast) is negotiated on a per endpoint ***(or per conference, some suggest)*** basis~~-media basis~~. ~~Note that control may also potentially be distributed/multicast.~~ This negotiation is performed per media, in the sense that, for example, audio/video may be distributed, while data and control are centralized.

Audio and video are sent using RTP via a non-guaranteed delivery means to minimize delay. Error concealment or other recovery action must be applied to overcome lost packets; in general audio/video packets are not re-transmitted since this would result in excessive delay in the LAN environment.[6] The assumption is that bit errors are detected in the lower layers, and errored packets are not sent up to H.22Z. Note that audio/video and call setup/control are never sent on the same LAN channel~~connection~~, and do not share a common PDU structure. Audio and video are sent to separate LAN ports using separate instances of RTP to allow for media-specific frame sequence numbers.

T.120 capabilities are negotiated via the H.22Z~~45~~ control LAN port using H.245, and upon receipt of appropriate messages, T.125 sessions are established on LAN using the LAN stacks of T.123 as appropriate. There is no synchronization of T.120 data with the audio/video in RTP, or with the ~~Q.931/H.245~~ control LAN port. T.120 is conveyed over the LAN between endpoints, or between the endpoint and the gateway on a fourth LAN connection (T.120 may used up to 4 LAN ports, one for each MCS priority). Thus, the typical point-to-point, or point-to-gateway link can be expected to have at least five~~three~~ (and up to nine~~seven~~ if MLP is used) LAN ports.

1. ~~1.~~ Audio/RTP
2. Audio/RTCP
3. ~~2.~~ Video/RTP
4. Video/RTCP
5. ~~3.~~ Control(H.22Z~~Q.931~~/H.245)
6. ~~4.~~ Data (T.120)[Up to 4 LAN ports]

Note that a well-known non-guaranteed delivery port is used for initial call setup. Since this port may be associated with many calls if it is associated with a GateKeeper or Gateway, it is not listed above. However, if it is associated with a terminal, then it is logically part of the above list.

[6] This is not to imply that retransmission is never used; restransmission of video in MacroBlocks is an important error recovery strategy. However, a simple reliance on traditional retransmission protocols is not sufficient due to the requirement for low delay operation.

~~*{Editor's Note: I have received more than one input that this lmitation is not needed since the higher layers have little control over LAN packet size in any case. It there is not disagreement, this rule will be removed}*~~

~~*{Editor's Notes: At the Haninge meeting, it was agreed to split H.22Z into two levels, and a list of needed functions was provided. I have included this list below, along with some comments on how they might (or might not) be met by RTP.*~~

~~*The functions of H.22Z are split into two levels. The lower level contains the following functions:*~~

- ~~*Timestamp used to measure quality of service. These are numbered across all media types. INPUT FROM IETF SUGGESTS THAT WE CAN USE THE JITTER FEATURE OF RTP FOR THIS PURPOSE.*~~
- ~~*PDU structure that allows the association of audio, video, data, and control as being from a given endpoint. RTP ALLOWS ASSOCIATION VIA SOURCE IDENTIFIER; ASSOCIATION IS ALSO KNOWN SINCE ALL LAN PORTS ARE ESTABLISHED BASED ON INITIAL CONTROL LAN PORT SIGNALING. DATA IS NOT COORDINATED IN TIME, BUT IS ASSOCIATED AT THE CONTROL LEVEL.*~~
- ~~*PDU structure that allows the determination of media type. DETERMINED IN PART FROM LAN PORT USED*~~

~~*The higher layer includes the following functions:*~~

- ~~*Media capture timestamps used to restore media clocks and to associate audio and video. IN RTP.*~~
- ~~*Any media specific messages such as freeze picture release. IN RTP?*~~
- ~~*Specific rules for packetizing each type of media. IN RTP, T.123.*~~
- ~~*Methods for recovering from lost packets; the methods vary with media type. IN RTP; SOME ADDITIONAL METHODS ADDED HERE.*~~
- ~~*Per media sequence numbers used to detect lost packets. USE RTP SEQUENCE NUMBERS ON SEPARATE LAN PORTS.*~~

~~*I have included this list as a reminder of what was agreed, and will remove it once it is apparent that RTP is satisfactory to all.}*~~

In some cases signals carried by the BAS channel in H.221 are carried with the media stream in H.22Z. This is described below in section XXX, and also in H.323. ***{Editor's Note: It is not clear that this is needed: discussion?}***

In general, two types of conference modes of operation on the LAN side are possible: distributed and centralized. It is also possible that different choices may be made for different media, e.g. distributed audio/video and centralized data. Procedures for determining what sort of conference to establish are in H.323; the procedures of H.22Z are intended to support all allowed combinations, according to the following:. ~~Thus, H.22Z provides for both point-to-point and~~

~~multi-cast links per media, with the exception of T.120 data, for which only centralized operation is described; with distributed T.120 being for further study.~~

1. Audio and video shall receive the same treatment, i.e. either both centralized or both disitributed.
2. Distributed control is for further study.
3. Distributed data is for further study.

6.2. Use of RTP/RTCP

A separate LAN port will be established for audio and video. On each LAN port a separate instance of RTP will be used, one for audio and one for video. Additional LAN ports are needed for RTCP, one for each RTP LAN port. Using H.245 signaling, additional audio and video channels (2 LAN ports per channel, one for RTP and one for RTCP) may be established if the terminal supports this capability.

In this section two documents, RTP and RTP profile (RTP-P for brevity) will be referred to(see references section for more information).

Unless an exception is specifically mentioned here, implementations shall follow those of RTP/RTP-P.

It should be kept in mind that the gateway may be acting as an RTP mixer, or it may be representing a single endpoint, or it may be representing many H.320 endpoints as one RTP source. The Gateway may also act as an RTP monitor

~~The timestamp field of RTP for both audio and video packets should be accurate to within 1/64,000.~~
~~{Editor's Notes: I have received input that the RTP accuracy is too tight for many PC implemenatations. Comments on the 1/64,000 value??? As I understand it, this goal is easily met in UNIX systems, but difficult on PCs. Hence the use of "should."}~~ {Current input is that RTP accuracy is acceptable. Comments}

The H.323 LAN terminal, when engaged in a any~~gateway-mediated~~ conference, whether point-to-point or multi-point, shall restrict the total video plus data bit rate averaged over a 5 second ***{Editor's Note: I have received a suggestion that this be left to the manufacturer to set at any rate they desire. Comments?}*** period ~~(in other words, keep the bursts to a minimum!),~~ to that signaled in the H.245 FlowControlCommands, H.245 logical channel commands, and the T.120 flow control mechanism.~~exchange. Thus, even though there may a very large "limit" on the bit rate that can be transmitted on the LAN, the sender shall ensure that the sum of audio, video, and data on the LAN side does not exceed that sum of audio, video, and data on the WAN side as signaled by the gateway~~. ~~The exceptions are~~

- ~~46/56 kbit/sec G.711, which is sent at the 64 kbit/sec rate on the LAN, and truncated or expanded in the H.323 gateway.~~
- ~~Possibly G.722~~
- ~~When the gateway is using a rate reducer; in this case the LAN side H.323 terminal shall matched the signaled rate, which will probably be less than the rate being sent over the WAN.~~

When the H.323 LAN terminal is connected to an H.323 gateway, the gateway shall use the means of H.245 and T.120 to force the H.323 terminal to be less than or equal to the WAN side media rates, with the following exceptions:

- Control bandwidth on the LAN does not match that in H.221.
- Audio bandwidth on the LAN may match that in H.221 on the WAN, but with gateway transcoding a match is not required.
- When the gateway is using a rate reducer; in this case the LAN side H.323 terminal shall match the signaled rate, which will probably be less than the rate being sent over the WAN.

6.2.1. Audio

Version (V): RTP-P does not mention (apparently) which version number is to be used; perhaps it is intended to be implied. In any case, we should use (2) ***{Editor's Note: This will be removed as soon as we confirm that the change has been made in RTP-P to refer to version 2 of RTP. An issue with RTP may also exist in that it would be desirable to have a larger set of version numbers so that we can refer to a fixed point in time more easily. Comments, especially from IETFers?}.***

CSRC Count(CC): This shall be set to zero for the H.323 endpoint and for the gateway when the H.323 gateway is not performing audio mixing. When the H.323 gateway supports centralized audio mixing, the CSRC information may be used to provide information to the terminal concerning what terminals are present in the audio sum. ***{Editor's note:*** ***~~What should the gateway set it to?? This seems problematic, as only 16 contributing sources are allowed, which is a lot less than the maximum H.231 conference size. Also, this seems like an awkward method for indicating who the loudest talker if is the gateway is acting as a mixer. One possibility is to require that the Gateway set CSRC to zero as well, and rely on H.245/H.243 to indicate conference information. Also, there is no function in H.243 that requires that the endpoint determine who the loudest talker is, or be aware of the current number of endpoints in the audio mix. The use of this field for distrubuted style conferences on the LAN requires further consideration.~~*** ***This seems less problematic now than it did a couple of months ago. Is provision of CSRC mandatory in RTP?? Should we make it mandatory here?}***

Payload Type(PT): Only ITU-T payload types such as (0)[PCMU], (8)[PCMA}, (9)[G722], and (15)[G728] shall be used. ***{Editor's Note: It is our hope that codepoints will be allocated for G.723 and G.dsvd so that we can make use of them. However, if this is not that case, we will rely on H.245 signaling and ignore the RTP payload type.}***

~~{New Editors Note: It appears that the SSRC serves the valuable function of allowing different program streams from one source to be distinquished. It isespecially important if a terminal is listening to more than one conference, where the H.243 terminal number will no longer suffice. Thus there seems no reason to modify the IETF rules for its useage.~~

~~However, H.323 terminals will still get and sent their terminal numbers via the H.245 control link, and should regard the SSRC as more of a program id than a terminal id~~
{Editor's Note: ***~~RTP-P requires a default 20 msec packet, with a requirement that the terminal be able to receive packets from 0 to 200 msec of audio. Is this acceptable to all? Should we tighten up on this in some fashion? I have received input that the 0-200 msec requirement is not compatible with much currently used PC hardware. Thus, this appears to be one area WHERE H.22Z WILL CLEARLY DEVIATE FROM RTP. The current suggestion is that the endpoint should signal via caps the packet size they can handle~~***

~~using H.245. We will need to define a required size range - suggestions? 50-150 msec mandatory???~~ I have received new input that RTP as is is acceptable. Comments?}

If the underlying LAN indicates a bad packet, or if a interruption in sequence numbers is observed, the receiver shall repeat the most recent octet ~~with a time constant~~ such that the value of the repeated octet decays to silence in 10 msec. ~~The choice of the time constant is left to the manufacturer.~~

When sending 48/56 kbit/sec PCM, the H.323 gateway shall pad the extra 1 or 2 bits in each octet, and use the RTP values for PCMA or PCMU(8 or 0). For Mu-law the padding consists of "1" in both the 7th and 8th bit. For A-law the 7th bit shall be 0 and the 8th bit 1.In the reverse direction the H.323 gateway shall truncate 64 kbit/sec G.711 on the LAN side to fit the G.711 rate being used in H.320. ***{Editor's note: Is there any need for the H.323 terminal to know that PCM 48/56 is being used? Another approach is to add codepoints to RTP, which seems feasible}***

{Editor's note: We need to confirm that RTP can handle 48, 56, and 64 Kbps G.722 as well; there appears as for G.711, to be only one code point. The same sort of issues arise}

When using RTCP, both RR and SR packets should be sent periodically as described in RTP. Only the CNAME SDES PDU should be used, and instead of the canonical name mentioned in RTP, the H.243 terminal identifier shall be sent. CNAME is used to associate audio and video for purposes of lip synchronization. **{Editor's note: Note that with more than one audio channel open, there is no H.245 level signal to associate that signal with a particular video logical channel, right???, so the CNAME is essential.}** Note that each Audio/Video pair must have a unique CNAME, and the procedure of RTP section 6.4.1 is not adequate to this task. The H.323 terminal shall append a conference id to the <M><T> pair to make it unique. The conference id is assigned (by the terminal? by the MCU? by the gateway?). ***~~{Editors note: We don't need CNAME since our SSRC is assured to be unique, but RTP says that CNAME is required in RTCP. Any other suggestions???}~~ {This issue requires more thought!}***

If possible, the H.323 termain should make use of the silence suppression feature of RTP, especially when the conference is multicast.

6.2.2. Video PDUs

Version (V): RTP-P does not mention (apparently) which version number is to be used; perhaps it is intended to be implied. In any case, we should use (2). ***{ See earlier editor's note.}***

CSRC Count(CC): This shall be set to zero for the H.323 endpoint and the gateway.
Payload Type(PT): Only ITU-T payload types such as that for H.261 shall be used. ***{Editor's Note: It is our hope that the IETF will include H.263 as well.}***

When using RTCP, both RR and SR packets should be sent periodically as described in RTP. Only the CNAME SDES PDU shall be used, and when the gateway is involved in the call instead of the canonical name mentioned in RTP, the H.243 terminal identifier shall be sent as the CNAME. On a point-to-point call CNAME shall follow the conventions of RTP. ~~encoded using ASCII? T.61?~~. ***{New Note: See Audio discussion of CNAME issues}~~{Editors note: Since we are now potentially receiving from multiple sources in more than one conference, this is once again important. The H.243 terminal id, although unique in a conference is not~~***

~~unique across conferences, so we may need to use the RTP type domain name, or some other endpint unique name, such as a room number, etc. }~~

~~{Editor's note: The issue has been raised that MB packetization is not always desirable. Further views on this topic are solicited; noting that MB is what RTP calls for, I would suggest this as a requirement, and allow other options, but I believe that those concerned want to avoid having MB packetization be required. Comments???}~~

H.261 MacroBlocks are the mandatory basis of video packetization as described in RTP-H261-REF.

6.2.3.Data PDUs

There are no special data PDUs; T.120 is used on the LAN as per T.123. Centralized vs distributed data conferencing on the LAN is described in H.323, and is negotiated via H.245. T.120 flow control on the LAN is managed using LAN protocoles, and not H.245 FlowControlCommands.

6.2.4. Overall Call Management using RTP/RTCP

{Editor's note: this section is probably not needed, but I am trying to clarify this in my own mind.}

In the simplist case of a point-to-point connection on the LAN, the SSRC is used to sychronization the received audio/video from terminal A, and the two streams are associated by a CNAME as specified in RTP.

When a gateway become involved in the call (either for a LAN/WAN point-to-point call, or a multi-point call involving one or more LAN endpoints), the H.323 gateway uses H.245 to assign H.243 terminal identifiers of the form <M><T> to each terminal on the LAN side. The <M><T> pair is unique within the conference, and can thus be used instead of the RTP CNAME.

In the event that a terminal sources more than one audio/video pair, a new CNAME is needed. This requires that something be added to the <M><T> pair to make it unique. If each pair is directed toward a separate conference, a conference id would be helpful, but if both audio/video pairs are directed toward the same conference, the termainl must append number to each <M><T>pair to make it unique.

In the event that a terminal participates in more than one conference at a time, it may potentially be receiving two pairs (or more) of audio/video that have the same <M><T> pair being used as the CNAME, since <M><T> pairs are only unique within a conference. The addition of a conference id to the <M><T> pair may have the effect of making the CNAME unique within the conference, but the terminal should be able to use its knowledge of H.22Z/H.245 signaling to distinguish the two pairs of audio/video since they will have separate, unique signaling channels.

It is a logical extension to suggest that these procedures apply to H.323 calls even if a gateway is not involved: RTP does not deal with having more than one CNAME per terminal.

In the event that a number of H.320 terminals are on the LAN side, e.g. A & B, where B is the broadcaster, and a terminal C is on the LAN side, it is unclear how the RTP timestamp can be used to provide lip sychronization. One solution might be to use the CSRC count, with the count fixed to 1. Thus, the SSRC in the CSRC field would be the source that the receiving endpoint should attempt to do lip synchronization with since this SSRC would correspond to the current broaccaster. ***Does this make sense???***

7. Control and Call Setup PDU Definitions

This section concerns the definition of PDUs for call setup, call control, and communications between the gateway, the gatekeeper, the e~~V~~MC(essential Multipoint~~Virtual MCU~~ Controller), and the H.323 terminal. Procedures for the use of these PDUs are in H.323. ***{Editor's Note: MC is too close to MCU. I suggest eMC to avoid confusion}***

7.1. Control PDU Structure

{Editor's Note: There is a suggestion that we don't need this: we will try to resolve at the next meeting}

The control PDU structure is as follows:

		Octet Number
Control Type (7 bits)	E	1
Message Body		2-N

Control Type: This 7 bit field has the values:

0	Reserved.
1	Q.931 Message
2	H.245 Message
3	H.22Z message
4	Reserved
5	Reserved for Q.2931 Messages
6-127	Reserved

E bit: When set to 1, indicates additional octets after the normal header. For this version of H.22Z, the E bit is zero.

~~***{Editor's note: We have agreed that the use of a gatekeeper on a call is optional. However, it may be that support of terminal/gatekeeper messages should be required for all terminals. Consider the issue.}***~~

Gatekeeper messages shall be supported for H.323 terminals and gateways. However, the user is not required to operate the H.323 terminals in a mode where authorization from the gatekeeper is needed for a new call.

7.2. Terminal and Gateway Registration PDUs

{Editor's Notes: There have been many suggestions for new messages; I will update the section after our discussion next week.}

7.2.1. Registration Request - RRQ

Note: Need to add way of signaling non-standard features during registration. ***{Editor's Note: Shouldn't non-standard features be signaled call-by-call?? However, I strongly agree with the direction proposed here on this point}***

Field	Field Size
Source Network Address	BYTE(16)
Source Identifier	CHAR(25)
Source Identifier Number	WORD
Year of Specification	CHAR(4) e.g. "1995"

Where ***{Editor's Note: this will be moved to Annex A since it is LAN specific}:***

- Network address is defined for each network as follows:

For IP:

Field	Field Size
Network and Host Address	BYTE(4)
Socket Number	BYTE(2)
Subnet Mask	BYTE(4)
Reserved	BYTE(6)

For IPX:

Field	Field Size
Network Address	BYTE(4)
Node Address	BYTE(6)
Socket Number	BYTE(2)
Reserved	BYTE(4)

For Other Networks:

For Further Study.

Note: Subnet Mask is required for IP networks to locate terminals within the network topology.

- Source Identifier is an optional text field for terminal or Gateway identification.

- Source Identifier Number is an optional field to allow a terminal to associate a number with its address for identification purposes:

7.2.2. Registration Confirmation - RCF

Field	Field Size

Capability Definition	TBD
Source Identifier	CHAR(25)
Source Identifier Number	WORD
Year of Specification	CHAR(4)

7.2.3. Registration Rejection - RRJ

Field	Field Size
Rejection Reason	WORD
Year of Specification	CHAR(4)

Where:

- Rejection Reason is defined as:

Rejection Reason	Code
Invalid Network Address	1
Duplicate Network Address	2
Others	?

7.3. Gatekeeper to Gatekeeper Configuration Messages

Note that how a terminal obtains a directory address is beyond the scope of H.22Z/H.323.

7.3.1. Sub-Domain List Request

This message can be used only for gatekeeper to gatekeeper operation.

Field	Field Size
Terminal Type	BYTE
Source Network Address	BYTE(16)

- Terminal Type is defined as:

Terminal Type	Code
H.323 Terminal	0
Gateway	1
Gatekeeper	2

7.3.2. Sub-domain Registered List

Field	Field Size
Number of Registered Terminals	WORD (16 bits)
Registered Terminal Record	Terminal Record(Number of Registered Terminals)

Where:

- Terminal Records are defined as:

Terminal Record	Code
Terminal Network Address	NETWORK ADDRESS

Terminal Type	BYTE

7.3.3. Gatekeeper Termination Indication

No Parameters

7.4. Connection Request Messages

7.4.1. Connection Request - CRQ

Note: Need to add means of addressing multiple nodes.

This message is sent to a well known port on that does not use guaranteed delivery means of operation.

Field	Field Size
Dynamic Control and Data Network Address (guaranteed link)	Network Address
Call Type	BYTE
Destination Network Address	Network Address
Bandwidth	WORD

Where:

- Call Type is defined as:

Call Type	Code
Point to Point Call	1
Multicast call	2

- Bandwidth is defined as the maximum amount of bandwidth required for the call in Kbits/second (total bandwidth in both directions). This could be used by the Gatekeeper in its connection approval determination processing. Note that Bandwidth is needed here, but is repated in the Q.931 messages.

{Editor's Note: Additional material will be added here on the usage of Q.931 messages for call progress, connection, etc. The messages will not be needed for a LAN to LAN connection, but will be needed if the gateway is involved in the call. }

7.5. Disconnect Request PDU

The purpose of these messages is to inform the gatekeeper that a call is dropping; they are sent only in the optional gatekeeper mode of operation. A mandatory Q.931 message will also be sent to the gateway.

7.5.1. Disconnect Request - DRQ

Field	Field Size

Disconnect Reason	BYTE

Where:

- Disconnect Reason is defined as:

Disconnect Reason	Code
Hang Up	1
Remote Site Aborted Call	2
LAN Bandwidth Limit Exceeded	3

Note: The Disconnect reason between peers is usually "Hang Up", but in the event of a power failure or problem at one end of a link, the reason code when the disconnect is sent to the Gatekeeper will contain "Remote Site Aborted Call"

7.5.2. Disconnect Confirmation - DCF

No Parameters.

7.6. Status Request Messages

7.6.1. Status Request - SRQ

No Parameters.

7.6.2. Status Response - SRR

Field	Field Size
Terminal Type	BYTE
Terminal Status	Status based on Terminal Type

- **For H.323 Terminals:**

Field	Field Size	Field Description
Call State	WORD	State of the call.
Packets Sent	DWORD	Number of total packets sent
Packets Received	DWORD	Number of total packets received
Packets Missed	DWORD	Number of total packets missed
Packets Received Out of Order	DWORD	Number of total packets received out of order

- **For Gateways:**

Field	Field Size	Field Description
Call State	WORD	State of the call.
Packets Sent	DWORD	Number of total packets sent
Packets Received	DWORD	Number of total packets received
Packets Missed	DWORD	Number of total packets missed
Packets Received Out of Order	DWORD	Number of total packets received out of order
Total Channel Rate	WORD	Audio Video and Data
Audio Receive Mode	WORD	H.320 audio receive mode
Audio Transmit Mode	WORD	H.320 audio transmit mode
Video Receive Mode	WORD	H.320 video receive mode

Video Transmit Mode	WORD	H.320 video transmit mode
Data Receive Mode	WORD	Data receive mode
Data Transmit Mode	WORD	Data transmit mode
Data Receive Rate	WORD	
Data Transmit Rate	WORD	
Multipoint Connect	WORD	

7.7. QOS Related PDUs

7.7.1. Adaptive Busy Indication (ABI)

Not clear that any particular content is needed.

8. Mechanisms for maintaining QOS

8.1.General Approach and Assumptions

At the call start, QOS may be signaled for as part of the initial CRQ PDU(proposed). The gatekeeper may grant or refuse the request. The following procedures do not assume that any particular QOS level has been supplied, and they are intended to provide the means for dealing with varying QOS levels.

The Sender and Receiver Reports of RTCP shall be the means by which QOS will be assessed. Logically, there are two types of congestion related delay that might be measured:

- Short term increases in delay that will result in a perceptible but not annoying slowing of the frame rate. These bursts are less than 1 ***{Editor's note: what is the right figure; this is too short}*** second in duration.
- A general rise in delay due to LAN congestion over time such that a feedback based mechanism is useful. This rise in congestion is measured over 1 minute intervals***{Editor's note: is this the right interval?}***Essentially, short term bursts are approached by error concealment, and a longer term congestion is approached by reducing the multi-media load. The assumption is made that all LAN multimedia terminals are H.323 terminals, and all will attempt to reduce LAN usage as congestion rises rather than "steal" bandwidth from each other.

8.2. Use of RTCP in Measuring QOS

8.2.1. Sender Reports

~~***{Editor's Note: It is recognized that this section needs more work}***~~The sender report servers three main purposes:

1. Allow synchronization of multiple RTP streams, such as audio and video.
2. Allow the receiver to know the expected data rate and packet rate.
3. Allow the receiver to measure the distance in time to the sender.

Of these three purposes, (1) is the most relevant to H.22Z. Manufacturers may make use of the sender reports in other ways at their discretion.

The relevant field for ~~RTP~~ stream synchronization is the RTP timestamp and the NTP timesamp in the sender report of RTCP. The NTP timestamp~~, which~~ gives "wall clock" time and corresponds to the RTP timestamp which has ~~with~~ the same units and random offset as the RTP capture timestamp in the media packets. The CNAME in RTCP binds the different SSRC identifiers from the same sender and can be used for synchronization. ~~Since both audio and video come from the same source, any inaccuracy or drift in the terminal "wall clock" is not important.~~ ~~***{Add more details on how this is used}***~~

~~When the difference between the audio and the video RTP timestamp exceeds 100 msec, indicating that audio and video are at least 100 msec out-of-synchronization, the receiver shall take recovery action by sending a ??? request to the sender. This message causes the sender to restart audio and video transmission in a synchronized fashion.~~ ~~***{Editor's note: At Haninge,***~~

~~*it was agreed that a procedure of this sort was needed. Does this still seem to be the case? Is this the right value? The right procedure?}*~~

8.2.2. Receiver Reports

Three parts of the Receiver Reports are used in H.22Z to measure QOS:

1. Fraction Lost
2. The cumulative packets lost
3. The extended highest sequence number received
4. Interarrival jitter

Items 2 and 3 are used to compute the number of packets lost since the previous receiver report. This can be taken as a long term measure of LAN congestion. See RTP section 6.3.4 for a sample computation. If this loss rate exceeds 1% ***{Editor's note: What is a good value?}*** the H.22Z terminal should reduce the media rates on the LAN side according to the procedures in section XXX below. If item 1 exceeds 1%, it may also be desirable to take corrective action. If the interval between receiver reports exceeds 5 minutes, H.323 terminals shall use item 1 as an indicator of serious congestion requiring media rate reduction on the LAN side.

Item 4 should be used as an indication of impending congestion. If interarrival increases for three consecutive receiver reports, the H.323 sending terminal should take corrective action.

{Editor's note: There is a strong element of guesswork in these numbers and rules. I have attempted to provide a kind of framework that will support interoperablity in behavior. Comments are especially welcome from those with experience in this area.}

8.3. ~~GLI~~ QOS Metrics

{Editor's Note: In this section I was working towards some specification of the general relationship between total LAN bandwidth, LAN fundamental bit rate, desired media delay parameters, and suggested packet size for audio and video. The obvious notion is that packet size can increase as LAN bit rate increases. However, it may be that this sort of thing should be left to the manufacturer, as it will vary from LAN to LAN. Still, it seems that we need to have some kind of limit on packet size for audio and video. RTP suggests 20 msec audio packets, with 0-200 msec range support required. What are your views?}

{I propose we wait on revising this section pending the QOS signaling proposal}

The following metrics are used to characterize the general LAN:

Maximum Bit Rate(MBR): The bit rate at which packets are sent on the LAN. For example, 10 Mbps for IEEE 802.3 is a common rate.

Instantaneous Packet Delay(IPD): The delay on a packet due to any arbitration or collisions that may take place. This delay varies from moment to moment, and often increases with load on the LAN.

Packet Jitter(PJ): The variation in IPD from packet to packet that must be compensated for by buffering on the receive side. Generally, the higher the jitter, the more delay across the LAN.

Media Bandwidth: On the LAN, the MBW is a product of the Packet size time the Packets/sec. For example, 64 kbit/sec PCM might be sent as 80 packets/sec of 100 bytes each.

Minimum Media Delay: On the LAN, the minimum delay that will be experienced is that needed to fill a packet plus the time needed for the packet to transit the LAN. Thus, to fill a 100 octet packet with PCM requires 12.5 msec. On a 10 Mbps LAN, sending the 800 bit packet requires about 80 usec. Whether we must double the packing delay to get overall delay depends on whether a checksum is computed on each packet

Actual Packet Delay: This is the sum of MMD(Minimum Media Delay) plus IPD(Instantaneous Packet Delay) plus Packet jitter (PJ). Note that jitter may raise or lower the total delay.

8.4. Procedures for Maintaining QOS

.A number of methods exist for the H.323 gateway/terminal to respond to an increase in packet loss or interarrival jitter in the far end receiver. These methods can be grouped into those that are appropriate for a rapid response to a short term problem, such as a lost or delayed packet, and those that are appropriate for a response to a longer term problem such as growing congestion on the LAN.

Short term responses:

- Reducing the frame rate for a short period of time. This may result in the H.323 gateway sending additional H.261 fill frames in the LAN->WAN direction to compensate for the packet underflow.

- Need to add additional short term responses for H.261

- ***{Editor's Note: A suggestion has been made that reducing the packet rate is the best way to deal with LAN congestion. Should we specify some method for providing larger and larger video packets/audio packets as congestion rises? The major problem I see with this approach is that it converts congestion/packet loss into delay. The other methods listed here seek to keep delay constant or nearly so, while decreasing picture quality or overall functionality. Your input??}***

Longer term responses:

- Reducing media bit rate(e.g. switching from 384 kbit/sec to 256 kbit/sec). This may involve a simple instruction to the encoder in a terminal, or it may involve the use of a rate reducer function in the H.323 gateway.

- Turning off media of lessor importance (e.g. turning off video to allow a large amount of T.120 traffic).

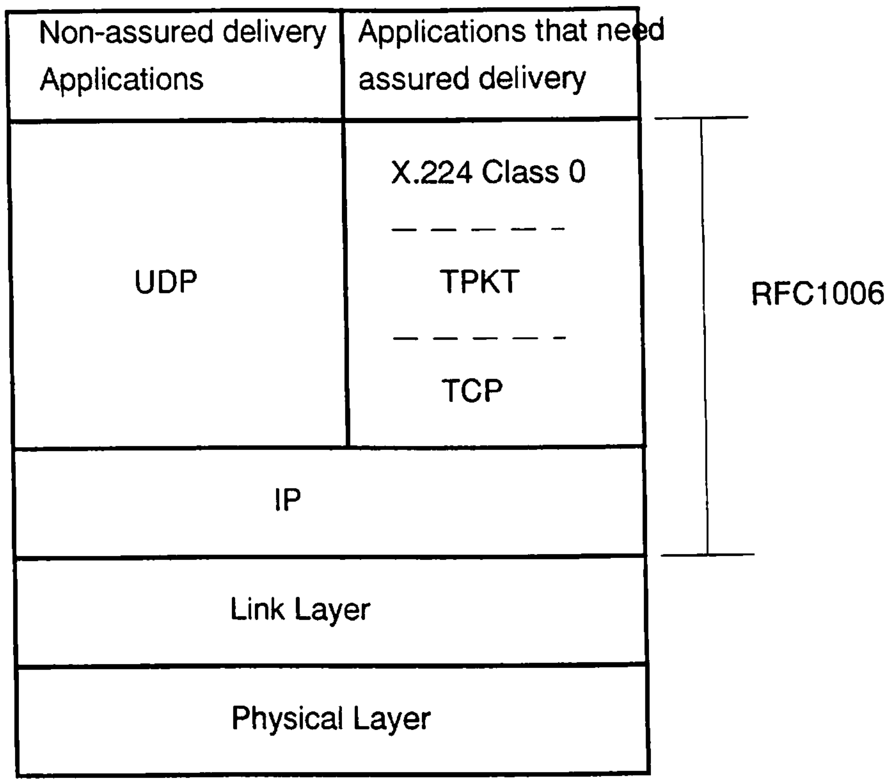
- Returning a busy signal (adaptive busy) to the receiver as an indication of LAN congestion. This may be combined with turning off a media, or even all media other than the control LAN port.

{Editor's Note: Should these procedures be detailed here, or in H.323? Clearly H.22Z must at least supply the adaptive busy related messages, but the media control messages already exist in H.245}

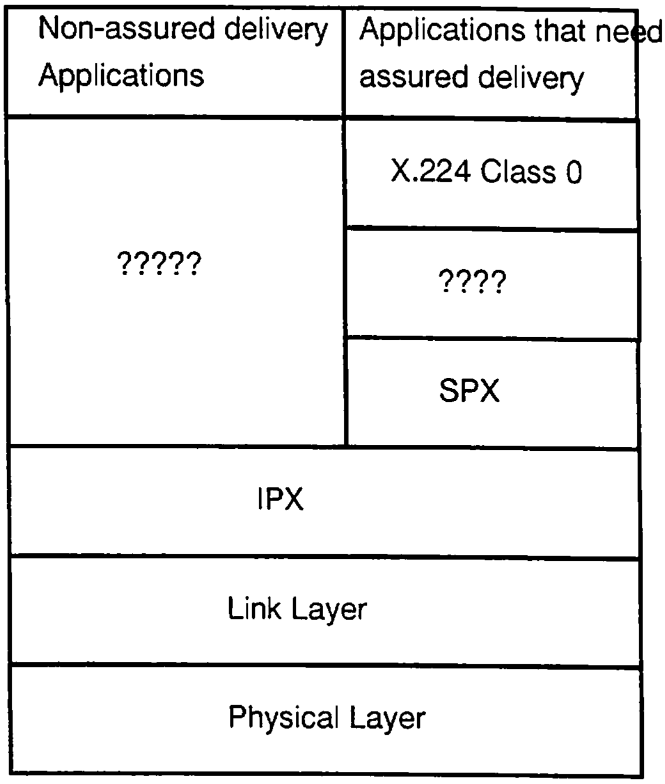
9. Annex A

This annex provides additional details concerning the operation of H.22Z on various actual LAN protocol stacks. This annex is non-normative. ***{Editor's Note: We may split the annex into normative and non-normative parts}***

9.1. TCP/IP/UDP



9.2. SPX/IPX


Non-assured delivery Applications
Applications that need assured delivery
X.224 Class 0
?????
????
SPX
IPX
Link Layer
Physical Layer