ETSI TS 103 161-3 V1.1.1 (2011-04)

Technical Specification

Access, Terminals, Transmission and Multiplexing (ATTM); Integrated Broadband Cable and Television Networks; IPCablecom 1.5; Part 3: Audio Codec Requirements for the Provision of Bi-Directional Audio Service over Cable Television Networks using Cable Modems

Reference
DTS/ATTM-003011-3

Keywords
access, broadband, cable, IP, multimedia, PSTN

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Important notice

Individual copies of the present document can be downloaded from:
http://www.etsi.org

The present document may be made available in more than one electronic version or in print. In any case of existing or perceived difference in contents between such versions, the reference version is the Portable Document Format (PDF). In case of dispute, the reference shall be the printing on ETSI printers of the PDF version kept on a specific network drive within ETSI Secretariat.

Users of the present document should be aware that the document may be subject to revision or change of status. Information on the current status of this and other ETSI documents is available at
http://portal.etsi.org/tb/status/status.asp

If you find errors in the present document, please send your comment to one of the following services:
http://portal.etsi.org/chaircor/ETSI_support.asp

Contents

Intellectual Property Rights

IPRs essential or potentially essential to the present document may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (http://webapp.etsi.org/IPR/home.asp).

Pursuant to the ETSI IPR Policy, no investigation, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Foreword

This Technical Specification (TS) has been produced by ETSI Technical Committee Access, Terminals, Transmission and Multiplexing (ATTM).

The present document is part 3 of a multi-part deliverable covering Access, Terminals, Transmission and Multiplexing (ATTM); Integrated Broadband Cable and Television Networks; IPCablecom 1.5, as identified below:

Part 1: "Overview";

Part 2: "Architectural framework for the delivery of time critical services over cable Television networks using cable modems";

Part 3: "Audio Codec Requirements for the Provision of Bi-Directional Audio Service over Cable Television Networks using Cable Modems";

Part 4: "Network Call Signalling Protocol";

Part 5: "Dynamic Quality of Service for the Provision of Real Time Services over Cable Television Networks using Cable Modems";

Part 6: "Event Message Specification";

Part 7: "Media Terminal Adapter (MTA) Management Information Base (MIB)";

Part 8: "Network Call Signalling (NCS) MIB Requirements";

Part 9: "Security";

Part 10: "Management Information Base (MIB) Framework";

Part 11: "Media Terminal Adapter (MTA) device provisioning";

Part 12: "Management Event Mechanism";

Part 13: "Trunking Gateway Control Protocol - MGCP option";

Part 14: "Embedded MTA Analog Interface and Powering Specification";

Part 15: "Analog Trunking for PBX Specification";

Part 16: "Signalling for Call Management Server";

Part 17: "CMS Subscriber Provisioning Specification";

Part 18: "Media Terminal Adapter Extension MIB";

Part 19: "IPCablecom Audio Server Protocol Specification - MGCP option";

Part 20: "Management Event MIB Specification";

Part 21: "Signalling Extension MIB Specification".

NOTE 1: Additional parts may be proposed and will be added to the list in future versions.

NOTE 2: The choice of a multi-part format for this deliverable is to facilitate maintenance and future enhancements.

1 Scope

The present document specifies the media aspects of the interfaces between IPCablecom client devices for audio and video communication. Specifically, it identifies the audio and video codecs necessary to provide the highest quality and the most resource-efficient service delivery to the customer. The present document also specifies the performance required in client devices to support future IPCablecom codecs, describes a suggested methodology for optimal network support for codecs.

The present document also extends the existing IPCablecom 1.0 Codec specification by introducing two new low-bit codecs, ITU-T Recommendation T.38 [20] fax relay for reliable fax transmission, RFC 2833 [24] DTMF Relay for reliable DTMF transmission and metrics to measure voice quality.

2 References

References are either specific (identified by date of publication and/or edition number or version number) or non-specific. For specific references, only the cited version applies. For non-specific references, the latest version of the reference document (including any amendments) applies.

Referenced documents which are not found to be publicly available in the expected location might be found at http://docbox.etsi.org/Reference.

NOTE: While any hyperlinks included in this clause were valid at the time of publication ETSI cannot guarantee their long term validity.

2.1 Normative references

The following referenced documents are necessary for the application of the present document.

[1] PKT-TR-ARCH1.5-V02-070412: "PacketCable 1.5, Architecture Framework Technical Report", April 12, 2007, Cable Television Laboratories, Inc.

[2] PKT-SP-DQOS1.5-I04-090624: "PacketCable 1.5 Specifications, Dynamic Quality-of-Service", Cable Television Laboratories, Inc.

[3] PKT-SP-NCS1.5-I03-070412: "PacketCable 1.5 Specifications, Network-Based Call Signaling Protocol", April 12, 2007, Cable Television Laboratories, Inc.

[4] ATIS-0152100-2005(R2010): "Packet Loss Concealment for Use with ITU-T Recommendation G.711".

NOTE: Available at http://webstore.ansi.org/RecordDetail.aspx?sku=ATIS-0100521.2005(R2010).

[5] Voice-Over-IP Forum Service Interoperability Implementation Agreement 1.0, December 1, 1997.

[6] "Current Methods of Speech Coding." R.V. Cox. International Journal of High Speed Electronics & Systems, Vol 8, No 1 (1997) pp 13-68.

[7] IETF RFC 1890 (1996): "RTP Profile for Audio and Video Conferences with Minimal Control".

[8] IETF RFC 2327 (1998): "SDP: Session Description Protocol".

[9] Void.

[10] Void.

[11] Void.

[12] Void.

[13] Void.

[14] Void.

[15] Void.

[16] IETF RFC 1889 (1996): "RTP: A Transport Protocol for Real-Time Applications".

[17] ITU-T Recommendation G.711 (1998): "Pulse code modulation (PCM) of voice frequencies".

[18] ITU-T Recommendation G.165 (1993): "Echo Cancellers".

[19] ITU-T Recommendation G.168 (2002): "Digital network echo cancellers".

[20] ITU-T Recommendation T.38 (2004): "Procedures for real-time Group 3 facsimile communication over IP networks".

[21] ITU-T Recommendation V.18 (2000): "Operational and interworking requirements for DCEs operating in the text telephone mode".

[22] ITU-T Recommendation G.728 (1992): "Coding of speech at 16 kbit/s using low-delay code excited linear prediction".

[23] ITU-T Recommendation G.729 (1998) Annex E: "11.8 kbit/s CS-ACELP speech coding algorithm".

[24] IETF RFC 2833 (2000): "RTP Payload for DTMF Digits, Telephony Tones and Telephony Signals".

[25] Telcordia GR-506 (2006): "LSSGR: Signaling for Analog Interfaces".

[26] Void.

[27] Void.

[28] Void.

[29] IETF RFC 3611 (2003): "RTP Control Protocol Extended Reports (RTCP XR)".

[30] ITU-T Recommendation P.56 (1993): "Objective measurement of active speech level".

[31] ITU-T Recommendation P.561 (1996): " In-service non-intrusive measurement device - Voice service measurements".

[32] ITU-T Recommendation G.107 (2003): "The E-Model: a computational model for use in transmission planning".

[33] ITU-T Recommendation P.862 (2001): " Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs".

[34] IETF RFC 3952 (2004): "Real-time Transport Protocol (RTP) Payload Format for internet Low Bit Rate Codec (iLBC) Speech".

[35] IETF RFC 3951 (2004): "Internet Low Bit Rate Codec (iLBC)".

[36] ANSI/SCTE 24-22 (2007): "iLBCv2.0 Speech Codec Specification for Voice over IP Applications in Cable Telephony".

NOTE: Available at http://www.scte.org/documents/pdf/Standards/ANSI_SCTE%2024-22%202007.pdf.

[37] IETF RFC 4298 (2005): "RTP Payload Format for BroadVoice Speech Codecs".

[38] ANSI/SCTE 24-21 (2006): "BV16 Speech Codec Specification for Voice over IP Applications in Cable Telephony".

[39] ITU-T Recommendation V.152 (2005): " Procedures for supporting voice-band data over IP networks".

[40] IETF RFC 2198 (1997): "RTP Payload for Redundant Audio Data".

[41] IETF RFC 3407 (2002): "Session Description Protocol (SDP) Simple Capability Declaration".

[42] ITU-T Recommendation G.722 (1988): "7 kHz audio-coding within 64 kbit/s".

[43] ITU-T Recommendation G.722 Appendix III: "A high quality packet loss concealment algorithm for G.722".

[44] ITU-T Recommendation G.722 Appendix IV: "A low-complexity algorithm for packet loss concealment with G.722".

[45] ITU-T Recommendation G.711 Appendix II (2000): "A comfort noise payload definition for ITU-T G.711 use in packet-based multimedia communication systems".

[46] ITU-T Recommendation V.150.1: "Modem-over-IP networks: Procedures for the end-to-end connection of V-series DCEs".

[47] ITU-T Recommendation V.22: "1200 bits per second duplex modem standardized for use in the general switched telephone network and on point-to-point 2-wire leased telephone-type circuits".

2.2 Informative references

The following referenced documents are not necessary for the application of the present document but they assist the user with regard to a particular subject area.

[i.1] PKT-SP-SEC1.5-I03-090624, PacketCable 1.5 Security Specification, June 24, 2009, Cable Television Laboratories, Inc.

[i.2] Void.

[i.3] Void.

[i.4] Void.

[i.5] Void.

[i.6] Void.

[i.7] Void.

[i.8] ITU-T Recommendation H.245 (2003): "Control protocol for multimedia communication".

[i.9] ITU-T Recommendation H.261 (1993): "Video codec for audiovisual services at p x 64 kbit/s".

[i.10] ITU-T Recommendation H.263 (1998): "Video coding for low bit rate communication".

[i.11] ITU-T Recommendation H.323 (2003): "Packet-based multimedia communications systems".

[i.12] ITU-T Recommendation H.324 (2002): "Terminal for low bit-rate multimedia communication".

[i.13] ITU-T Recommendation Q.24: "Multifrequency push-button signal reception".

3 Definitions and abbreviations

3.1 Definitions

For the purposes of the present document, the following terms and definitions apply:

audio server: audio server plays informational announcements in IPCablecom network

NOTE: Media announcements are needed for communications that do not complete and to provide enhanced information services to the user. The component parts of audio server services are media players and media player controllers.

call management server (CMS): controls the audio connections

NOTE: Also called a call agent in MGCP/SGCP terminology. This is one example of an application server.

cable modem termination system (CMTS): device at a cable head-end which implements the DOCSIS RFI MAC protocol and connects to CMs over an HFC network

delay: absolute time required for a signal to transit from source to receiver

dynamic quality of service (DQoS): assigned on the fly for each communication depending on the QoS requested

hybrid fibre/coaxial cable (HFC): HFC system is a broadband bidirectional shared media transmission system using fibre trunks between the head-end and the fibre nodes, and coaxial distribution from the fibre nodes to the customer locations

Internet control message protocol (ICMP): extension to the Internet Protocol, ICMP supports packets containing error, control and information messages

jitter: variability in the delay of a stream of incoming packets making up a flow such as a voice communication

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

media gateway (MG): provides the bearer circuit interfaces to the PSTN and transcodes the media stream

media gateway controller (MGC): overall controller function of the PSTN gateway

NOTE: Receives, controls and mediates call-signalling information between the IPCablecom and PSTN.

multimedia terminal adapter (MTA): contains the interface to a physical voice device, a network interface, codecs, and all signalling and encapsulation functions required for VoIP transport, class features signalling and QoS signalling

off-net call: communication connecting an IPCablecom subscriber out to a user on the PSTN

on-net call: communication placed by one customer to another customer entirely on the IPCablecom network

pulse code modulation (PCM): commonly employed algorithm to digitize an analog signal (such as a human voice) into a digital bit stream using simple analog to digital conversion techniques

quality of service (QoS): guarantees network bandwidth and availability for applications

registered Jack-11 (RJ-11): standard 4-pin modular connector commonly used for connecting a phone unit into a wall jack

real-time transport protocol (RTP): protocol for encapsulating encoded voice and video streams

transit delays: time difference between the instant at which the first bit of a PDU crosses one designated boundary, and the instant at which the last bit of the same PDU crosses a second designated boundary

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

user datagram protocol (UDP): connectionless protocol built upon Internet protocol (IP)

NOTE: Delay and latency are similar concepts and frequently used interchangeably. However, delay focuses on the time to transit from transmitter (such as a speaker's mouth) to a receiver (such as a listener's ear), while latency focuses on the time to transit from a receiver to a transmitter, as would be the case for a signal going through a piece of equipment.

3.2 Abbreviations

For the purposes of the present document, the following abbreviations apply:

ANS Answer Tone

NOTE: As per ITU-T Recommendation V.25.

ASCII American Standard Code for Information Interchange
CDMA Code Division Multiple Access
CED Facsimile CallED tone

NOTE: Defined in ITU-T Recommendation T.30.

CIF Common Intermediate Format
CM DOCSIS Cable Modem
CMS Call Management Server
CMTS Cable Modem Termination System
CNG Facsimile Calling tone

NOTE: Per ITU-T Recommendation T.30.

Codec Coder-DECoder
CPE Consumer Premise Equipment
CPM Continuous Presence Multipoint
CSRC Contributing source lists
DIS Digital Identification Signal
DOCSIS® Data-Over-Cable Service Interface Specification
DQoS Dynamic Quality of Service
DSC Dynamic Service Change
DTMF Dual-tone Multi Frequency (tones)
FEC Forward Error Correction
GOB Group of Blocks
GSM Global System for Mobility
HFC Hybrid Fibre/Coaxial cable
ICMP Internet Control Message Protocol
INTRA intra
IP Internet Protocol
ISDN Integrated Services Digital Network
ISP Internet Service Provider
IVR Interactive Voice Response System
LCO Local Connection Option
LSSGR LATA Switching System Generic Requirements
LUB Least-Upper-Bound
MG Media Gateway
MGC Media Gateway Controller
MIB Management Information Base
MOS Mean Opinion Score
MOS-CQ Mean Opinion Score-Conversational Quality
MOS-LQ Mean Opinion Score-Listening Quality
MTA Multimedia Terminal Adapter
NCS Network Call Signalling
NTSC National Television Standards Committee

NOTE: Defines the analog colour television, broadcast standard used today in North America.

PAL Phase Alternate Line

NOTE: The European colour television format that evolved from the American NTSC standard.

PCM Pulse Code Modulation
PCMA Pulse Code Modulation A-Law

NOTE: As defined in ITU-T Recommendation G.711 [17].

PCMU Pulse Code Modulation μ-law

NOTE: As defined in ITU-T Recommendation G.711 [17].

PDU Protocol Data Unit
PSTN Public Switched Telephone Network
QCIF Quarter Common Intermediate Format
QoS Quality of Service
RAM Random Access Memory
RJ-11 Registered Jack-11
RR Receiver Report
RR Receiver Report
RSVP Resource Reservation Protocol
RTCP Real-time Transport Control Protocol
RTP Real-time Transport Protocol
SDP Session Description Protocol
SQCIF Sub-Quarter Common Intermediate Format
SR Sender Report
SR Sender Report
SSRC Synchronization Source

NOTE: Telephony, real-time control protocol.

TCP Transmission Control Protocol
TDD Telecom Devices for the Deaf
TDM Time Division Multiplex(ing)
TDM Time Division Multiplexing
TDMA Time Division Multiplexing Access
TTY Text Telephone
UDP User Datagram Protocol
UDPTL UDP Transport Layer

NOTE: A transport protocol defined in ITU-T Recommendation T.38 [20].

VAD Voice Activity Detection
VBD Voice-band Data
VBR Variable Bit Rate
VoIP Voice over IP
XR Extended Reports

4 Void

5 Background

This clause outlines the IPCablecom 1.5 architecture support elements and the DOCSIS network infrastructure necessary to deliver quality audio and video service. It is intended to clarify external interfaces and functional requirements necessary to implement the targeted audio and video quality using speech and video codecs.

The key requirement for voice communications using IP transmission is the ability to attain "toll" or better audio quality. Given the variable nature of shared packet mediums and the stringent human-factor requirements of this quality standard, it is necessary to optimize multiple system parameters to attain this goal. Additionally, IPCablecom has been tasked with offering superior quality, exceeding current PSTN standards where feasible. Key requirements from the IPCablecom product definition requiring architectural optimization for codecs follow.

5.1 IPCablecom Voice Communications Quality Requirements

As defined in the IPCablecom architecture document [1], requirements for toll-quality voice communications service in IPCablecom include numerous metrics to ensure competitive or superior quality and service to the PSTN. In order to support these requirements, network plant and equipment may have to be groomed. In order to provide guidelines for that grooming, several network implications affecting codec performance are discussed below.

5.2 Network Preparation for Codec Support

The critical areas of network performance, which must be optimized in tandem with codecs, are packet loss, latency, and jitter. Elaboration of network/codec implications for each of these areas follows.

5.2.1 Packet Loss Control

There is a direct correlation between packet integrity and audio quality. Anecdotal codec research suggests initial 3 % packet loss rate results, on average, in a reduction in Mean Opinion Score (MOS) scores of 0,5 point, on a scale of 5. Due to less-than-pristine conditions and human-detectable compromises with most codecs, the resulting audio quality for a 3 % packet loss rate will be well below PSTN "toll" quality. Above 3 %, codec performance falls off rapidly, and resulting voice quality is unacceptable.

Applications and/or codecs may provide error correction or concealment mechanisms, which may increase latency through buffering. Once latency thresholds have been exceeded, the tradeoff between latency and fidelity becomes an untenable situation.

5.2.2 Latency Control

Control of overall latency requires a hand-in-hand effort by the system resources and the application-in this case, a speech or video application dominated by the codec component.

There are multiple device elements and network components inducing latency during traversal of an audio signal from capture of the speaker's voice until reception at the receiver's ear. The primary contributors to latency for an on-net voice and off-net communication along this path are:

- Audio sampling and analog-to-digital conversion.
- Buffering of samples (audio framing, plus look-ahead).
- Compression processing.
- Packetization of compressed data.
- Local network (DOCSIS) traversal.
- Routing to the backbone network.
- Backbone traversal.
- Far-end reception of packets and traversal of local access.
- Buffering of out-of-order and delayed packets.
- Decoding, decompression, and reconstruction of the audio stream.

The major contributors to codec-related latency in the network are described below.

5.2.2.1 Latency Control: Buffering

While network jitter and corresponding buffering increase call latency, another source of buffering can be induced by the application as a corrective response to severe packet loss. Although the ultimate solution to additional buffering delay is a pristine network, realistically some packet loss will occur.

Accounting for lost packets suggests the need for support concealment or reconstruction of lost data, and in many instances these techniques employ some mechanism of redundant information encoding, temporally shifting and embedding audio frames in the data stream. This not only increases the effective bandwidth requirement, but also creates, in effect, an additional buffer to allow for reassembly, increasing latency.

In order to apply certain reconstruction methodologies in an optimal fashion, the application needs accurate data regarding the statistical characteristics of the media stream. Some information is available through real-time control protocol (RTCP) mechanisms, such as a gross measure of packet loss. Additional information, such as burst frequency and predictive time-of-day effects, would improve the potential of the application to make optimal adjustments. Planning for the collection and analysis of this type of network information will allow developers more options in the future, potentially creating applications that will increase network utilization efficiency or quality.

5.2.2.2 Latency Control: Optimal Framing/Packetization

As outlined in clause 5.2.1, the loss of audio data frames can have a severe impact on audio quality. The packing of multiple audio frames into a single packet will exacerbate the problem, effectively expanding the loss of one packet into the loss of multiple adjacent audio frames of data. This also increases latency by buffering larger portions of audio samples prior to sending.

One way to minimize these effects is to send small packets containing the minimum number of frames. This will increase bandwidth use by increasing the header-to-data ratio for packets, but will minimize latency and potentially increase reconstruction quality. This suggests that the optimal packet size for voice applications is fairly small, containing compressed information for one, two, or, at most, three frames of sampling data (typically corresponding to 10, 20, or 30 milliseconds of voice frames).

5.2.2.3 Latency Control: Packet Timing Optimization

To avoid additional buffering delay, packets shall be sent at a rate equal to integral multiples of the audio sample frame rate of the codec. This synchronization results in lockstep between the codec framing and packet transmission.

The frame sizes of the codecs are shown in table 1. Default packetization periods are specified in [7].

Table 1: Frame Sizes of the Codecs

Codec	Frame Size (msec)
G.711	0,125
iLBC	20
iLBC	30
BV16	5
G.728	0,625
G.729E	10
G.722	0,0625

5.2.3 Codec Transcoding Minimization

Transcoding occurs whenever a packetized voice signal encounters an edge device without compatible codec support. Transcoding introduces additional latency during the decode/recode stage. Additionally, if transcoding resources at the edge gateway are shared, additional delay can be introduced.

Transcoding between compressed codecs also results in degradation of the original sample, as current codec compression techniques are not loss less. In the event that a combination of transcoding and packet loss causes a signal to be reduced below minimum quality, it is likely that a higher bandwidth codec will be employed. Thus, transcoding artifacts can result in the unintended side effect of higher system bandwidth utilization.

In the case of on-net and off-net IP connections, transcoding can be eliminated if all necessary codecs are supported on the client. This is, in fact, impractical but can be optimized statistically if a device supports multiple codecs and can be updated periodically.

5.2.4 Bandwidth Minimization

There are two primary mechanisms that client devices may employ to minimize the amount of bandwidth used for their audio/video applications:

- A compressed, low bitrate codec may be applied, thus reducing the bandwidth required.
- A codec may employ some form of variable bitrate transmission.

The selection of codecs occurs at the device's discretion or via network selection, depending on the protocol employed. Regardless, this takes place after the initial capabilities exchange to determine a compatible codec between endpoints, and assumes that the requested bandwidth is granted by the bandwidth broker element.

Variable rate transmission may occur when a codec employs methods resulting in a non-constant bitstream representation of voice data. Voice activity detection (VAD) - silence suppression - is a basic form of variable rate transmission, sending little or no data during speaker silence periods. More advanced variable bitrate encoding (VBR) occurs when a codec dynamically optimizes the compression bitstream.

6 Device Requirements for Audio codec support

As markets evolve, endpoint codecs will change too, and neither a provider nor a customer can be expected to replace their cable modem/MTA frequently to accommodate these market changes. Given the rapid growth of the digital wireless market in particular, it is likely that, at some point, a statistically significant portion of voice communications will require a new codec in the standard suite in order to maintain voice quality.

Since interconnection between diverse codecs requires transcoding - which introduces unwelcome latency and artifacts - one goal of the IPCablecom network is to minimize transcoding. Thus, a forward-looking approach to codec evolution is necessary - one which supports the most important interconnect codecs, as well as improved performance of on-net codecs introduced in the marketplace over the next several years.

However, now and for the immediate future, it is not cost-feasible to provide support for every possible interconnecting codec. Thus, a compromise must be established limiting the required power of the processors and local memory. Therefore, IPCablecom requires a minimum threshold of programmable upgradeability in its MTA devices, as described below. These requirements include support for downloading new software from an authorized system resource, headroom in processing for slightly more complex new codecs, and additional local storage to hold program data.

6.1 Dynamic Update Capability

All MTA devices shall be capable of downloading new software from authorized sources.

6.2 Maximum Service Outage

If the MTA supports life-line services (such as 911 emergency service), service disruption shall not exceed 20 seconds excluding reboot time when downloading new software to the MTA.

6.3 Minimum Processing Capability

All MTA devices shall be capable of supporting the equivalent simultaneous execution of codec combinations shown in the following table. Although the present specification does not mandate the support of either G.728 or G.729 Annex E, this requirement provides the necessary reserve capacity for additional future codecs to be provisioned (configured and downloaded) on the MTA. The MTA shall support T.38 fax relay on all ports simultaneously. Media Gateway shall be configurable to allow a specified proportion of ports to transmit T.38 fax simultaneously. However, the use of T.38 fax relay and a voice codec on a given port for both the MTA and Media Gateway is mutually exclusive at any given time.

In addition DTMF Relay and Voice Metrics shall be supported on all connections simultaneously by both MTA and Media Gateway.

Table 2: MTA Processing Capability

Maximum Ports supported by MTA	G.711 ports	iLBC Ports	BV16 Ports	G.728 ports	G.729E ports
1	1				
1		1			
1			1		
1				1	
1					1
2	2				
2		2			
2			2		
2				2	
2					2
2	1	1			
2	1		1		
2	1			1	
2	1				1
3	3				
3	2	1			
3	2		1		
3	2			1	
3	2				1
3	1	2			
3	1		2		
3	1			2	
3	1				2
4	4				
4	3	1			
4	3		1		
4	3			1	
4	3				1
4	2	2			
4	2		2		
4	2			2	
4	2				2
More than 4	For future study	For future study	For future study	For future study	For future study

6.4 Minimum Audio Codec Storage Capability

All MTA devices shall be capable of maintaining simultaneously, in device memory or storage, all mandatory and recommended codecs specified herein (i.e. equivalent storage for G.711, G.728, G.729 Annex E, internet Low Bit rate Codec [iLBC™], and BroadVoice™16 [BV16]). Although the present specification does not mandate either G.728 or G.729 Annex E, this requirement provides reserve capacity for additional codecs to be provisioned to the MTA in the future.

Although it is necessary to provide storage for all mandatory and recommended codecs, the minimum run-time memory only needs to support one of the recommended codecs along with G.711, iLBC and BV16, subject to the minimum processing specification in clause 6.3.

7 Audio Codecs Specifications

7.1 Feature Support

Offering a competitive and/or superior product requires support for more than toll-quality delivery of audio. In addition to features and signalling capabilities, which are beyond the scope of the present document, the audio codec application must provide transparent support for certain audio features. These include general detection mechanisms, DTMF, fax, analog modem, echo compensation, and hearing-impaired support.

7.1.1 DTMF Support

Dual-tone multi-frequency (DTMF) support allows employment of dual-tone multiple frequency signals by either an autodialing system or through manual entry of tones. In order for DTMF tones to be captured correctly by the receiving device, tonal integrity (frequency accuracy and signal duration) must be maintained even through compression and transcoding.

IPCablecom endpoints (MTAs and MGs) shall successfully pass DTMF tone transmissions in band via RFC 2833 [24] telephone events (clause 7.1.9) subject to a successful negotiation. When negotiation is unsuccessful, e.g. due to interworking with older non-RFC2833-capable endpoints, DTMF tone transmissions shall be passed in the regular audio stream using the voice codec by MTAs and MGs.

The capability described above shall be supported on all connections.

7.1.2 Fax and Modem Support

IPCablecom needs to support analog fax and modem interfaces for two reasons. First, fax and modem equipment are common in residences, and customers will continue to use these familiar devices for some years to come. Second, even with cable modem access, many SOHO or ISP users will continue to access their dial-up networks using a traditional modem.

In order to provide customers with access for analog fax and modems, the MTA devices shall be able to detect fax/modem signals and signal these detections using the appropriate protocol. The codec at each end is then switched to G.711 for the remainder of the session. Additionally, echo cancellation is disabled in response to a disabling signal sent by some devices (fax or modem) consisting of a 2 100 Hz tone with periodic phase reversals per ITU-T Recommendations G.165 [18] and G.168 [19]. After the device session has completed, echo compensation shall be enabled.

A more robust solution for supporting fax is to employ fax relay. Fax relay involves demodulating the T.30 transmission and sending control and image data over the IP network. At the receiving end, the received data is remodulated and sent to the fax terminal using another T.30 session. This is described in the ITU-T standard T.38 [20]. MTAs and Media Gateways shall support T.38 fax relay as defined in clause 7.1.8.

MTAs and MGs shall detect the T.30 fax preamble (V.21 flags) and CNG (calling fax tone). The detection of CNG shall be a configurable option since it will cause calls between Super Group 3 fax machines to drop back to standard Group 3 rates (14,4 kb/s max) in T.38 implementations not capable of supporting Version 3 (V.34). If CNG detection is disabled, calls between Super Group 3 fax machines will be treated as modem calls (with transmission rates of up to 33,6 kb/s) as these devices do not send the T.30 fax preamble once they recognize each other through their V.8 handshaking at the start of the call. On the other hand, enabling CNG detection as a trigger to switchover to T.38 will ensure that all fax calls benefit from the use of fax relay to provide resilience from packet loss. MTAs and MGs detecting CNG shall apply appropriate signal discrimination to minimize the chance that a voice call could inadvertently be switched to T.38 fax relay.

A more robust solution for supporting modem and TTY is to employ voice band data transmission using the method described in ITU-T Recommendation V.152 [39]. V.152 involves quickly switching to a codec that can accurately relay modem and TTY signals over an IP network. The use of V.152 with RFC 2198 [40] redundancy, makes the transmission more resilient to packet loss in the network. This is an important feature for V.152 since packet loss causes modems and TTY to drop in speed or disconnect. MTAs and Media Gateways may support V.152 with RFC 2198 [40] redundancy as defined in this specification.

7.1.3 Echo Compensation Support

When end-to-end delay in an audio communication is more than 20 milliseconds, an artifact called line echo can occur. This echo, if not removed, will be heard by the remote talker (thus it is also called talker echo) whenever he or she speaks.

Line echo is created at the telephone interface of the MTA, or the PSTN interface of the PSTN gateway. A device called a hybrid coil (or hybrid) converts the separate audio transmit and receive signals (four-wire interface) into a single two-wire interface compatible with a standard telephone. This conversion by the hybrid creates an echo back to the remote talker. An echo canceller is used to remove this echo.

Line echo cancellation shall be provided in IPCablecom MTA and Gateway devices to mitigate the effects of line echo. This echo canceller shall allow both parties to speak simultaneously (double-talk), so that one talker does not seize the line and block out the other user from being heard.

The performance of the line echo canceller shall comply with either ITU-T Recommendations G.165 [18] and G.168 [19].

During periods when only the remote talker is speaking, the local echo canceller should either inject comfort noise or allow some noise to pass through to the remote talker, so that a "dead-line" is not perceived. However, if local voice activity detection (VAD) is enabled, either the noise injection should be disabled, or the echo canceller should communicate its state with the VAD, in order for the VAD to not estimate the injected noise mistakenly as the true background noise.

In an application where the MTA is located in a home, the length of the echo canceller is typically short (8 msec or less). For PSTN gateway applications, the echo canceller length is typically much longer (32 msec or longer). Vendors may choose to differentiate their products by providing longer echo canceller lengths suitable for their application, or other programmable parameters.

In MTAs where a non-standard telephone interface is used (e.g. four-wire microphone and headset) and the MTA has no hybrid coils, line echo cancellation may not be necessary. However, where a microphone and speakers are used, acoustic echo cancellation may be necessary, and vendors implementing these products should employ acoustic echo cancellation.

7.1.4 Asymmetrical Services Support

MTA devices should be capable of supporting employment of different codecs for upstream and downstream audio channels. This allows potential optimization of device resources, network bandwidth, and user service quality.

7.1.5 Hearing-impaired Services Support

For over one million hearing-impaired North Americans and 20 million North Americans with some amount of hearing loss, TTY (teletype technology) equipment can be the primary communication link to the outside world. This type of equipment has evolved lacking the type of standardization allowing broad interoperability among international manufacturers. The ITU, as recently as February 1998, adopted the ITU-T Recommendation V.18 [21] to begin alleviating this problem. Recommendation V.18 attempts to outline a procedure, which includes protocol negotiation, for connecting these devices.

Since CPE for the hearing impaired consists of text input/output devices coupled with voice-band modems, any system designed to support them would need to be able to pass DTMF and voice-band modem tones coherently. Typically, these devices will interface to the PSTN via an acoustical coupler to a phone or with a regular RJ-11 telephone jack.

MTA devices shall support detection of ITU-T Recommendation V.18 [21] hearing-impaired tones, including V.18 Annex A. Upon detection of a V.18 signal, the MTA shall notify the CMS of the Telecom Devices for the Deaf (TDD) Event, if this event is in the Requested Events list. When a terminating MTA detects answer tone from a TDD, the MTA shall notify the CMS of the modem tone event, if this event is in the Requested Events list. The MTA shall disable echo cancellation for the remainder of the session when phase reversals are present in the answer tone, in accordance with ITU-T Recommendation G.168 [19].

Upon detection of a V.18 signal, the codec at each end shall be switched to a codec that supports transmission of V.18 tones for the remainder of the session. These codecs are recommended: G.711, G.726 at 32 kbps, G.726 at 40 kbps. The endpoints shall change codecs at the direction of the CMS, unless multiple codecs have been negotiated between the endpoints when the connection was established. Depending upon the specific codecs negotiated for the connection, the endpoints shall reserve and/or commit additional HFC bandwidth to accommodate the requirements of the new codec.

7.1.6 A-law and µ-law Support

Both companding modes (µ-law and A-law) of G.711 shall be supported.

7.1.7 Packet Loss Concealment

All Media Gateways and Media Terminal Adaptors shall detect audio packet loss and implement some method to conceal losses from end-users. Specifications for low bit rate codecs (e.g. G.728, G.729, iLBC, BV16) include methods for concealment (the packet loss concealment method for iLBC, as defined and included in [34] is RECOMMENDED for iLBC and the packet loss concealment method for BV16, as defined and included in [38] is RECOMMENDED for BV16). For G.711, the method defined in ATIS-0152100-2005(R2010) [4] is RECOMMENDED. For G.722, the method defined in either [43] or [44] is RECOMMENDED.

7.1.8 Fax Relay

IPCablecom needs to support fax interfaces since fax equipment continues to be used by both residential and business customers. The recommended solution for supporting fax is to employ Call Management Server or Media Gateway Controller controlled fax relay. Fax relay involves demodulating the T.30 transmission and sending control and image data over the IP network. At the receiving end, the received data is remodulated and sent to the fax terminal using another T.30 session.

The ITU-T Recommendation T.38 is a widely recognized standard for fax relay [20]. The first version for the T.38 specification is version 0 and the majority of implementations are compatible with this version, while later implementations are also required to inter-operate with version 0. MTAs and Media Gateways shall support version 0 of the T.38 specification [20] in order to ensure interoperability with existing T.38 implementations. In addition, MTAs and Media Gateways may support versions 1 and 2 of T.38. MTAs and Media Gateways shall not use version 3. MTAs and Media Gateways shall support the V.27ter, V.29, V.17 modem protocols for page transmission within the T.38 implementation to allow transfer rates up to 14 400 bps. Fax transmissions utilizing the V.34 modem protocol (super G3 fax) should be handled as described in clause 7.1.2 using the G.711 pass-through mode. However, if CNG detection is enabled as a trigger for T.38, version 0, 1, or 2 shall be used to force a down-speed to Group 3 rates at a maximum of 14 400 bps [20]. MTAs and Media Gateways that are capable of T.38 version 3 (but have not negotiated it) shall set the V.8 Capabilities bit (bit 6) of the DIS frame to 0 if a DIS frame is received with the V.8 Capabilities bit set to 1. This locks the fax transmission to Group 3 rates by preventing a return to V.8 negotiations. This requirement applies to DIS frames received on both the packet and TDM interfaces of Media Gateways and on both the packet and analog interfaces of MTAs.

7.1.8.1 T.38 Over UDPTL

T.38 version 0 allows for a number of transport options including TCP and UDP. The UDP transport option is referred to as UDPTL in [20]. MTAs and Media Gateways shall support UDPTL. Within UDPTL, additional options allow support for redundancy or forward error correction. MTAs and Media Gateways shall support redundancy and may support FEC. When using redundancy, a redundancy level of 4 shall be used for T.30 control message data and a redundancy level of 1 shall be used for T.4 phase C data.

T.38 does not currently define any security authentication or privacy mechanisms for UDPTL; consequently T.38 sessions using UDPTL will not have secure media at the transport level.

T.38 Annex D describes the set of attributes to be used when setting up a T.38 UDPTL session. For more information on the use of these attributes refer to [3].

To control the T.38 UDPTL session, the FXR package will be used and all endpoints shall support this package as described in [3].

The MTA shall be prepared to receive a T.38 UDPTL fax packet of at least 160 bytes in the downstream. This is based on 40 ms packetization period and a 14 400 bps data rate. It includes the UDPTL datagram without the IP and UDP headers.

Upon transition to UDPTL T38, MTAs and media gateways should immediately send T.38 "No signal" indicator packets if the MTA or media gateway would not otherwise be sending signal or data packets. For DQoS considerations, T.38 fax packets should use the same port used by the voice packets for the connection. In addition, MTAs and media gateways shall send T.38 fax packets at a default 20 ms packetization period in the upstream unless directed by the CMS via the packetization period to use a different packet rate (10 ms / 20 ms / 30 ms).

Table 4 shows the DQoS flowspec parameters for 10 ms / 20 ms / 30 ms T.38 sessions (with redundancy of 1 for the T.4 data) that can be used in the least-upper-bound calculations for Authorization and resource requests. If the fax session is performed using the fxr/gw mode, then the data flow shall fit within the DQoS flow characteristics described above.

7.1.8.2 T.38 Over RTP

T.38 running over the RTP protocol as described in [20] is currently out of scope.

7.1.9 DTMF Relay

RFC 2833 [24] specifies in-band RTP payload formats and usage to carry DTMF, modem and fax tones, line states, and call progress tones across an IP network either as recognized "telephone events" or as a set of parameters defining a tone by its volume, frequency, modulation and duration of its components. Besides the transport of tones across an IP network, [24] also allows for the remote collection of DTMF digits by a media gateway to relieve an Internet end system (e.g. media server) of having to do this. Other advantages of [24] include inherent redundancy to cope with packet loss and the means to allow IP phones to generate DTMF digits when signalling to the PSTN without requiring DTMF senders.

The use of RTP payloads in RFC 2833 [24] to carry telephone events, states and telephony tones represents an in-band means of signal transmission as opposed to an out-of-band path via the CMS.

For DTMF, IPCablecom endpoints shall support transmission and reception of RFC 2833 [24] DTMF telephone-events 0-15 which represents the minimum level required for compliance with the RFC. IPCablecom endpoints may support other telephone-events. If negotiated for a call, these events shall be transferred via RFC 2833 telephony event packets regardless of the codec specified for the speech. In addition as an RTP payload type, DTMF relay shall be secured through the IPCablecom bearer encryption and authentication mechanisms defined in [i.1], if these are active on a call. MTAs and MGs shall support the mandatory security options listed in [i.1] for DTMF relay and additionally, if the optional encryption algorithms are supported for audio codecs, then these shall also be supported for DTMF relay.

RFC 2833 [24] references ITU-T Recommendation Q.24 [i.13] in defining the minimum DTMF tone duration of 40 ms. Additionally, ITU-T Recommendation Q.24 [i.13] includes a duration range lower than 40 ms when the DTMF tones may be accepted as DTMF digits (as low as 20 ms). For North American networks, Telcordia's LSSGR [25] specifies that tone durations greater than 40 ms must be accepted (subject to rise/fall times of less than 5 ms) and tones between 23 ms and 40 ms may be accepted by receivers. However generators should provide 50 ms minimum tone duration (with a rise/fall time < 3 ms). Receivers should accept minimum inter-digit times of 40 ms. Total on-off cycle times of 93 ms are to be accepted but 100 ms is to be generated as both minimum and objective.

RFC 2833 [24] does not specify DTMF tone duration requirements at the egress gateway instead relying on DTMF detection accuracy at the ingress gateway. Considering the industry requirements, IPCablecom endpoints shall detect DTMF tones of 40 ms or more and report their duration relative to the RTP timestamp. Endpoints may detect DTMF digits of duration greater than 23 ms but endpoints shall not report DTMF digits when their duration is less than 23 ms.

An IPCablecom endpoint shall not transmit a DTMF telephone-event packet containing a duration field of value zero. An IPCablecom endpoint should ignore a received DTMF telephone-event packet containing a duration field of value zero.

The repetition rate of RFC 2833 telephony event packets in the transmit direction shall be equal to the same packetization time as the selected audio codec. Therefore the repetition rate of RFC 2833 [24] packets has the same range as packetization intervals, i.e. 10 ms, 20 ms, and 30 ms.

In accordance with [24], unless a mutually exclusive event (detection of new DTMF digit) occurs, the final packet of each event shall be transmitted a total of three times at the specified packetization interval with the E-Bit flag set. This repetition will generally ensure satisfactory performance in the event of the occasional lost packet. However, if another DTMF digit is detected before the two redundant end-of-event packets are sent, the retransmission shall be aborted and instead the new DTMF telephone event reported using the regular packetization interval.

Upon receipt of any telephone-event packet, IPCablecom endpoints shall play out the tone on the Time Division Multiplexing (TDM) interface for the Media Gateways and Line Interface for the MTAs. Since the signal is received on the IP interface and not the TDM interface, this does not constitute a signalling event and the Call Agent or Media Gateway Controller shall not be informed of this.

RFC 2833 [24] describes two options for telephone event play out. Either the tone may be played out for the duration specified in the telephone event payload or it may be played out continuously until it is stopped when an end of event or mutually-exclusive event packet is received, an audio packet is received, or a timeout expires after a period with no packets. Because of its robustness against packet loss, IPCablecom endpoints shall use the continuous method of play out.

RFC 2833 [24] allows for the ingress media gateway to either replace the audio packets when transmitting telephone-event packets or send both audio and telephone events concurrently. To avoid increasing the bandwidth requirements in DQoS systems, an ingress media gateway shall stop sending audio and replace audio packets with RFC 2833 [24] DTMF telephone-event packets whenever a DTMF digit is detected. When replacing the audio, at the moment an event is detected the audio packet being constructed at the time of detection should be discarded.

DTMF telephone-events shall be fully played out by an egress gateway according to the duration specified in the event subject to an optional minimum play-out duration that may be provisioned on the endpoint. If audio data is also received by an egress gateway for the same timestamp period as covered by telephone-event packets, the egress gateway should overwrite the audio to the extent it remains in the play-out buffer. If some of the audio event has already played out due to a jitter buffer having adapted down to a low value, the telephone event play out may be shortened from the duration specified in the RFC 2833 [24] event but not below the minimum play-out duration as this would compromise the ability for a short duration DTMF tone to be detected when a low-bit-rate audio codec is in use. This is necessary even when the ingress (transmitting) gateway replaces the audio transmission when sending telephony-event packets, as there will still be some delay before this can take effect, i.e. the event recognition time. During this time nothing can prevent the telephony signal being transferred across the network and potentially played out from the egress gateway. When tone play-out by the egress gateway is per a minimum provisioned duration, the egress gateway shall enforce a 45 ms inter-digit time (silence) following play-out of the DTMF tone.

As already stated, the last telephone-event packet indicating the end of event will generally be transmitted 3 times. Audio packets being replaced by RFC 2833 [24] packets shall continue to be suppressed during the redundant transmission of the end-of-event packets.

7.1.10 V.152 Transmission

IPCablecom needs to support modem equipment interfaces since many residential and business customers still make use of dial-up modem lines for various services. These services include dial-up network access for work, home security systems, and many home electronic devices. In addition, support for TTY is also necessary to support hearing or speech impaired customers. The recommended solution is to support V.152 voice band data transmission along with RFC 2198 [40] redundancy. The combination of these technologies allows for modem and TTY signals to pass through an IP network reliably even when small amounts of packet loss exist. MTAs and Media Gateways shall support a redundancy level of 1 for V.152. MTAs and Media Gateways may support redundancy levels higher than 1 subject to QOS availability.

Table 4 shows the DQoS flowspec parameters for 10/20/30 ms voice band data sessions (with a redundancy level of 1 for usage of G.711 as a V.152 codec) that can be used in the least-upper-bound calculations for Authorization and resource requests.

7.1.10.1 V.152 Transition Triggers

The following tones are all triggers for switching to voice band data transmission: 1 100 Hz (CNG), V.21 preamble, V.18 Annex A, 2 100 Hz (for example ANS also known as CED, ANSam, /ANS, /ANSam). If an MTA or Media Gateway detects V.18 Annex A or 2 100 Hz tone, and V.152 has been negotiated for the connection, the endpoint shall transition to V.152 mode, 2 225 Hz answer tone as per ITU-T Recommendation V.150.1 [46] Appendix VI and Unscrambled binary ones signal as per ITU-T Recommendation V.22 [47].

MTAs and Media Gateways shall transition to V.152 mode on the receipt of packets that are the negotiated payload type for V.152 mode. This ensures that both ends will be switched into V.152 mode as soon as possible. MTAs and Media Gateways shall transition from V.152 mode to voice mode after detecting RTP packets that have non-VBD payload types, if V.152 mode VBD packets have been both sent and received in a session and the MTA or Media Gateway starts receiving RTP packets that have a previously negotiated non-VBD payload type.

Upon transition to V.152 mode, either through local detection of the tone on the analog interface or reception of a V.152 packet as described above, MTAs must perform the following steps (does not apply to Media Gateways):

1) MTAs shall initiate a DSC to the CMTS to request the additional bandwidth necessary for the V.152 payload including redundancy.

2) Once the extra resources are reserved and committed, MTAs shall begin to send V.152 packets with the negotiated payload type on the connection and play out any received data to the analog interface.

Given the sensitive nature of analog modems, MTAs should send RTP packets at the amount of bandwidth that is authorized for the existing voice session until the resources required for V.152 are committed. This allows tones to be played continuously to the receiving end during QoS reservation.

7.1.10.1.1 Considerations for Simultaneous T.38 and Voice Band Data Support

Certain signals that would cause a T.38 switch also cause a switch to voice band data. Since T.38 is more reliable and consumes less bandwidth than V.152 with redundancy, T.38 is the preferred method for transmitting fax calls.

If both T.38 and V.152 have been specified by the CMS, the CMS preference for T.38 and V.152 use must be followed when using t38-strict and t38-loose and V.152. The following requirements take into account whether V.152 has been negotiated between the endpoints and whether RFC 3407 [41] capability information indicating the necessary support for T.38 is present.

1) t38-loose/v152 specified and V.152 is not negotiated

   If t38-loose has been specified by the CMS as the preferred mode for fax, and V.152 has not been negotiated, T.30 fax preamble (V.21flags) shall cause the T.38 procedure to start and CNG may cause the T.38 procedure to start. V.152 will not be used for fax.

2) t38-loose/v152 specified and V.152 negotiated

   If t38-loose is specified by the CMS in the LCO as the preferred method of handling fax, and the two endpoints have negotiated V.152 use, the IPCablecom endpoint shall start the T.38 procedure if V.21preamble is detected. In addition, the IPCablecom endpoint may start the T.38 procedure event if CNG is detected and the endpoint is provisioned to use CNG as a T.38 trigger. If the endpoint is provisioned not to use CNG as a fax detection mechanism, then it shall enter V.152 mode upon CNG detection.

3) v152/t38-loose specified and V.152 negotiated

   If the LCO has specified that V.152 is preferred over t38-loose for fax handling and the two endpoints have negotiated V.152 use, the IPCablecom endpoint shall enter V.152 mode upon CNG or V.21preamble detection.

4) v152/t38-loose specified and V.152 not negotiated

   If the LCO has specified that V.152 is preferred over t38-loose for fax handling and the two endpoints have not negotiated V.152 use, T.30 fax preamble (V.21flags) shall cause the T.38 procedure to start and CNG may cause the T.38 procedure to start. V.152 will not be used for fax.

5) t38-strict/v.152 specified and V.152 negotiated and 3407 present

   If the LCO has specified that t38-strict is preferred over V.152 for fax handling and the two endpoints have negotiated V.152 use, the IPCablecom endpoint shall start the T.38 procedure if the rcd contains RFC 3407 [41] capability information indicating the necessary support for t38 AND V.21 preamble is detected. In addition, the IPCablecom endpoint may start the T.38 procedure if the rcd contains RFC 3407 [41] capability information indicating the necessary support for t38 AND CNG is detected and the endpoint is provisioned to use CNG as a T.38 trigger. If the endpoint is provisioned not to use CNG as a fax detection mechanism, then it shall enter V.152 mode upon CNG detection.

6) t38-strict/v.152 specified and v.152 negotiated, but 3407 not present

If the LCO has specified that t38-strict is preferred over V.152 for fax handling and the two endpoints have negotiated V.152 use, and the rcd does not contain RFC 3407 [41] capability information indicating the necessary support for t38, then the gateway shall enter V.152 mode upon CNG or V.21preamble detection.

7) v.152/t38-strict specified and V.152 negotiated (it doesn't matter if 3407 present or not)

If the LCO has specified that V.152 is preferred over t38-strict for fax handling and the two endpoints have negotiated V.152 use, the IPCablecom endpoint shall enter V.152 mode upon CNG or V.21preamble detection.

8) v.152/t38-strict specified and V.152 not negotiated and 3407 is present

If the LCO has specified that V.152 is preferred over T.38-strict for fax handling and the two endpoints have not negotiated V.152 use, the IPCablecom endpoint shall start the T.38 procedure if the rcd contains RFC 3407 [41] capability information indicating the necessary support for t38 AND V.21preamble is detected. In addition, the IPCablecom endpoint may start the T.38 procedure if the rcd contains RFC 3407 [41] capability information indicating the necessary support for t38 AND CNG is detected AND the endpoint is provisioned to use CNG as a T.38 trigger. If the endpoint is provisioned not to use CNG as a fax detection mechanism, then CNG shall be sent via a negotiated audio codec.

7.1.11 Security Considerations

V.152 uses RTP and RTCP for its transmission. The security requirements for RTP and RTCP defined in the security specifications [i.1] shall be followed.

7.2 Mandatory Codecs

The following codecs shall be supported in all MTAs and MGs.

7.2.1 G.711

G.711 (both µ-law and A-law versions) [17] shall be supported in all MTAs and MGs. This codec provides toll-quality voice and is ubiquitous. It provides the "fallback" position for services such as fax, modem, and hearing-impaired services support, as well as common gateway transcoding support. In addition, G.711 is used as the fallback mode if there are not enough resources to establish a new connection using the requested codec (e.g. two channels of the G.728 or G.729 Annex E are already in existence, and there are not enough resources for a third connection to use a compressed codec).

7.3 Recommended Codecs

In addition to G.711, it is RECOMMENDED that MTAs and MGs also support at least one of the following codecs.

7.3.1 iLBC

iLBC [34], [35] should be supported in all MTAs and MGs. IPCablecom has as a mandate to provide toll or superior voice quality. iLBC is a mid-bitrate (13,3 kb/s and 15,2 kb/s), high-quality solution. When iLBC is supported, both the 20 ms and 30 ms frame size modes shall be supported. iLBC provides high quality, low-bandwidth performance and high packet loss robustness for on-net calls and ensures high performance for applications such as IVR systems. It was created to provide a codec suitable for IP communication networks. In addition, it provides DTMF pass through. Experimental track IETF RFC "internet Low Bit Rate Codec (iLBC)" [35] contains the iLBC source code in floating point C.

A fixed point reference code implementation of iLBC is available for IPCablecom in [36] along with test vectors for verification of correct bit exact implementation. The fixed point code is provided to assist vendors in product development in order to ease implementation, testing and verification, and to guarantee quality.

7.3.2 BV16

BroadVoice16 [37], [38] should be supported in all MTAs and MGs. IPCablecom has as a mandate to provide toll or superior voice quality. BroadVoice16 (BV16) is a mid-bitrate, high-quality solution. BV16 provides high quality, low-bandwidth performance for on-net calls and ensures high performance for applications such as IVR systems. In addition, it provides DTMF pass through. It was created to provide a codec suitable for IP communication networks.

7.4 Optional Codecs

MTAs and MGs may support the following codecs.

7.4.1 G.728

G.728 [22] may be supported in all MTAs and MGs. IPCablecom has as a mandate to provide toll or superior voice quality. G.728 provides high quality, low-bandwidth performance for on-net calls and ensures the highest possible performance for applications such as IVR systems. In addition, it provides superior background noise handling, as well as medium quality music carriage.

7.4.2 G.729 Annex E

G.729 Annex E [23] may be supported in all MTAs. IPCablecom has as a mandate to provide toll or superior voice quality. G.729E is a mid-bitrate (11,8 kb/s), high-quality solution. G.729 Annex E provides high quality, low-bandwidth performance for on-net calls and ensures the highest possible performance for applications such as IVR systems. In addition, it provides superior background noise handling, as well as medium quality music carriage.

7.4.3 G.722

G.722 [42] is the earliest international standard on wideband speech coding. G.722 may be supported in wideband-capable MTAs and MGs. If G.722 is used, MTAs and MGs shall support 10 ms, 20 ms, and 30 ms packetization rates. G.722 is a multi-rate wideband speech codec for 16 kHz sampled signals. It has three selectable bit rates: 48 kb/s, 56 kb/s and 64 kb/s. The 48 kb/s version of G.722 produces medium-quality wideband speech, and the 56 kb/s and 64 kb/s versions produce good- to high-quality wideband speech. MTAs and MGs using the G.722 codec shall support 64 kb/s and may support 56 kb/s and 48 kb/s.

7.5 Optional Features

7.5.1 Wideband Codecs

Given that the majority of early customers will be "black phone" users, support for wideband (i.e. greater than circuit voice bandwidth) codecs on either MTAs or MGs is not being mandated. However, some vendors optionally may choose to differentiate their product by selecting components that will support higher fidelity in the event a wideband codec is provisioned through methods specified in clause 6.1. Furthermore, some IPCablecom applications may generate wideband media with their application-specific devices and without the involvement of MTAs or MGs.

7.5.2 Optional Codecs

A vendor may supply any codecs not described herein.

7.5.3 Voice Activity Detection (VAD)

A vendor may employ VAD to reduce bandwidth consumption. If employed, this capability shall be optional, allowing disabling. Some codecs have associated VAD implementations (e.g. G.729B), while many others do not (e.g. G.711, G.728, and G.722). In the latter cases, the VAD implementation shall adhere to the IMTC Voice-Over-IP Forum Service Interoperability Implementation Agreement 1.0 [5]. For use with the G.722 codec, MTAs and MGs should employ VAD and silence suppression (Discontinuous Transmission - DTX) to reduce bandwidth using a mechanism of the vendor's choice. If silence suppression is used with G.722, User Equipment and Media Gateways should transmit Silence Insertion Descriptor frames as specified in [45].

7.6 Session Description of Codecs

Session descriptor protocol (SDP) messages are used to describe multimedia sessions for the purposes of session announcement, session invitation, and other forms of multimedia session initiation. SDP descriptions are used in Network Call Signalling (NCS) [3]. This clause describes the required specification of the codec in SDP, and the required mapping of the SDP description into RSVP flowspecs.

A typical SDP description contains many fields that contain information regarding the session description (protocol version, session name, session attribute lines, etc.), the time description (time the session is active, etc.), and media description (media name and transport, media title, connection information, media attribute lines, etc.). The two critical components for specifying a codec in an SDP description are the media name and transport address (m) and the media attribute lines (a).

The media name and transport addresses (m) are of the form:

m=<media> <port> <transport> <fmt list>

The media attribute line(s) (a) are of the form:

a=<token>:<value>

A typical IP-delivered voice communication would be of the form:

m=audio 3456 RTP/AVP 0

a=ptime:10

On the transport address line (m), the first term defines the media type, which in the case of an IP voice communications session is audio. The second term defines the UDP port to which the media is sent (port 3456). The third term indicates that this stream is an RTP Audio/Video profile. Finally, the last term is the media payload type as defined in the RTP Audio/Video Profile, RFC 1890 [7]. In this case, the 0 represents a static payload type of μ-law PCM coded single channel audio sampled at 8 KHz. On the media attribute line (a), the first term defines the packet formation time (10 ms).

Payload types other than those defined in [7] are dynamically bound by using a dynamic payload type from the range 96-127, as defined in [8], and a media attribute line. For example, a typical SDP message for G.726 would be composed as follows:

m=audio 3456 RTP/AVP 96

a= rtpmap:96 G726-32/8000

The payload type 96 indicates that the payload type is locally defined for the duration of this session, and the following line indicates that payload type 96 is bound to the encoding "G726-32" with a clock rate of 8 000 samples/sec.

Codecs defined in this specification shall be encoded with the following string names in the rtpmap parameter.

Table 3: Codec RTP Map Parameters

Codec	Literal Codec Name	RTP Map Parameter
G.711 μ-law	PCMU	PCMU/8000
G.711 A-law	PCMA	PCMA/8000
iLBC	iLBC	iLBC/8000
BroadVoice16	BV16	BV16/8000
G.726 at 16 kb/s	G726-16	G726-16/8000
G.726 at 24 kb/s	G726-24	G726-24/8000
G.726 at 32 kb/s	G726-32	G726-32/8000
G.726 at 40 kb/s	G726-40	G726-40/8000
G.728	G728	G728/8000
G.729A	G729	G729/8000
G.729E	G729E	G729E/8000
RFC 2198 [40] Redundancy(for V.152 only)	red	red/8000
RFC 2833 [24] DTMF	telephone-event	telephone-event/8000
G.722 at 48 kb/s	G722-48	G722-48/8000
G.722 at 56 kb/s	G722-56	G722-56/8000
G.722 at 64 kb/s	G722-64	G722-64/8000

For use in the SDP, the rtpmap parameter (i.e. PCMU/8000 in the case of μ-law, or PCMA/8000 in the case of a-law) is used. Unknown rtpmap parameters should be ignored if they are received.

For every defined Codec (whether it is represented in SDP as a static or dynamic payload type), the following table describes the mapping that shall be used from either the payload type or ASCII string representation to the bandwidth requirements for that Codec.

It is important to note that the values in table 4 do not include any bandwidth that may be required for media security (authentication, 2 or 4 byte value as outlined in the security specification), and the actual values used in resource allocation may need to be adjusted to accommodate IPCablecom security considerations.

For non-well-known codecs, the bandwidth requirements cannot be determined by the media name and transport address (m) and the media attribute (a) lines alone. In this situation, the SDP must use the bandwidth parameter (b) line to specify its bandwidth requirements for the unknown codec. The bandwidth parameter line (b) is of the form:

b= <modifier> : <bandwidth-value>

For example:

b= AS:99

The bandwidth parameter (b) will include the necessary bandwidth overhead for the IP/UDP/RTP headers. In the specific case where multiple codecs are specified in the SDP, the bandwidth parameter should contain the least-upper-bound (LUB) of the desired codec bandwidths.

The mapping of RTP/AVP code to RSVP Flowspec (as used by Dynamic Quality of Service [2]) shall be according to table 4.

Table 4: Mapping of Session Description Parameters to RSVP Flowspec

Parameters from Session Description			Flowspec parameters		Comments
RTP/AVP code	**Rtpmap**	**Ptime (msec)**	**Values b,m,M[1]**	**Values r,p[2]**	
0	<none>	10	120 bytes	12 000 bytes/sec	G.711 µ-law using the Payload Type defined by IETF
0	<none>	20	200 bytes	10 000 bytes/sec	
0	<none>	30	280 bytes	9 334 bytes/sec	
96-127	PCMU/8000	10	120 bytes	12 000 bytes/sec	G.711 µ-law PCM, 64 kb/sec, default Codec
96-127	PCMU/8000	20	200 bytes	10 000 bytes/sec	
96-127	PCMU/8000	30	280 bytes	9 334 bytes/sec	
8	<none>	10	120 bytes	12 000 bytes/sec	G.711 A-law using the Payload Type defined by IETF
8	<none>	20	200 bytes	10 000 bytes/sec	
8	<none>	30	280 bytes	9 334 bytes/sec	
96-127	PCMA/8000	10	120 bytes	12 000 bytes/sec	G.711 A-law PCM, 64 kb/sec, default Codec
96-127	PCMA/8000	20	200 bytes	10 000 bytes/sec	
96-127	PCMA/8000	30	280 bytes	9 334 bytes/sec	
96-127	iLBC/8000	20	78 bytes	3 900 bytes/sec	iLBC, FB-LPC, 15,2 kb/s, 20 ms frame size with 5 ms lookahead; 13,3 kb/s, 30 ms frame with 10 ms lookahead
96-127	iLBC/8000	30	90 bytes	3 000 bytes/sec	
96-127	BV16/8000	10	60 bytes	6 000 bytes/sec	BV16 (narrow-band), 16 kb/sec
96-127	BV16/8000	20	80 bytes	4 000 bytes/sec	
96-127	BV16/8000	30	100 bytes	3 334 bytes/sec	
96-127	G726-16/8000	10	60 bytes	6 000 bytes/sec	
96-127	G726-16/8000	20	80 bytes	4 000 bytes/sec	
96-127	G726-16/8000	30	100 bytes	3 334 bytes/sec	
96-127	G726-24/8000	10	70 bytes	7 000 bytes/sec	
96-127	G726-24/8000	20	100 bytes	5 000 bytes/sec	
96-127	G726-24/8000	30	130 bytes	4 334 bytes/sec	
2	<none>	10	80 bytes	8 000 bytes/sec	G.726-32, identical to G.721, which is assigned Payload Type 2 by IETF
2	<none>	20	120 bytes	6 000 bytes/sec	
2	<none>	30	160 bytes	5 334 bytes/sec	
96-127	G726-32/8000	10	80 bytes	8 000 bytes/sec	
96-127	G726-32/8000	20	120 bytes	6 000 bytes/sec	
96-127	G726-32/8000	30	160 bytes	5 334 bytes/sec	
96-127	G726-40/8000	10	90 bytes	9 000 bytes/sec	
96-127	G726-40/8000	20	140 bytes	7 000 bytes/sec	
96-127	G726-40/8000	30	190 bytes	6 334 bytes/sec	
15	<none>	10	60 bytes	6 000 bytes/sec	G.728, assigned Payload Type 15 by IETF
15	<none>	20	80 bytes	4 000 bytes/sec	
15	<none>	30	100 bytes	3 334 bytes/sec	
96-127	G728/8000	10	60 bytes	6 000 bytes/sec	G.728, LD-CELP, 16 kb/s
96-127	G728/8000	20	80 bytes	4 000 bytes/sec	
96-127	G728/8000	30	100 bytes	3 334 bytes/sec	
18	<none>	10	50 bytes	5 000 bytes/sec	G.729A, identical to G.729, assigned Payload Type 18 by IETF
18	<none>	20	60 bytes	3 000 bytes/sec	
18	<none>	30	70 bytes	2 334 bytes/sec	
96-127	G729/8000	10	50 bytes	5 000 bytes/sec	G.729A, CS-ACELP, 8 kb/s, 10 ms frame size with 5 ms lookahead
96-127	G729/8000	20	60 bytes	3 000 bytes/sec	
96-127	G729/8000	30	70 bytes	2 334 bytes/sec	
96-127	G729E/8000	10	55 bytes	5 500 bytes/sec	G.729E, CS-ACELP, 11,8 kb/s, 10 ms frame size with 5 ms lookahead
96-127	G729E/8000	20	70 bytes	3 500 bytes/sec	
96-127	G729E/8000	30	85 bytes	2 834 bytes/sec	
96-127	red/8000	10	205 bytes	20 500 bytes/sec	RFC 2198 [40] Redundancy used for V.152 transmission only. These numbers are for the G.711 used as a V.152 codec with redundancy of level 1
96-127	red/8000	20	365 bytes	18 250 bytes/sec	
96-127	red/8000	30	525 bytes	17 500 bytes/sec	
N/A	N/A	10	80 bytes	8 000 bytes/sec	T.38 fax relay packets (with T.4 redundancy level 1, T30 redundancy level 4)
N/A	N/A	20	116 bytes	5 800 bytes/sec	
N/A	N/A	30	152 bytes	5 067 bytes/sec	
N/A	N/A	10	62 bytes	6 200 bytes/sec	T.38 fax relay packets

Parameters from Session Description			Flowspec parameters		Comments
RTP/AVP code	Rtpmap	Ptime (msec)	Values b,m,M[1]	Values r,p[2]	
N/A	N/A	20	80 bytes	4 000 bytes/sec	(without redundancy)
N/A	N/A	30	98 bytes	3 267 bytes/sec	
9	<none>	10	120 bytes	12 000 bytes/sec	G.722 at 64 kb/s using the Payload Type defined by IETF
9	<none>	20	200 bytes	10 000 bytes/sec	
9	<none>	30	280 bytes	9 334 bytes/sec	
96-127	G722-48/8000	10	100 bytes	10 000 bytes/sec	G.722 at 48 kb/s using dynamic payload type
96-127	G722-48/8000	20	160 bytes	8 000 bytes/sec	
96-127	G722-48/8000	30	220 bytes	7 334 bytes/sec	
96-127	G722-56/8000	10	110 bytes	11 000 bytes/sec	G.722 at 56 kb/s using dynamic payload type
96-127	G722-56/8000	20	180 bytes	9 000 bytes/sec	
96-127	G722-56/8000	30	250 bytes	8 334 bytes/sec	
96-127	G722-64/8000	10	120 bytes	12 000 bytes/sec	G.722 at 64 kb/s using dynamic payload type
96-127	G722-64/8000	20	200 bytes	10 000 bytes/sec	
96-127	G722-64/8000	30	280 bytes	9 334 bytes/sec	
NOTE 1: b is bucket depth (bytes). m is minimum policed unit (bytes). M is maximum datagram size (bytes). NOTE 2: r is bucket rate (bytes/sec). p is peak rate (bytes/sec).					

7.6.1 iLBC Session Description

Parameters are mapped to SDP in a standard way. When conveying information by SDP, the encoding name shall be "iLBC" (the same as the MIME subtype [35]).

If 20 ms frame size mode is used, local iLBC encoder shall send "mode" parameter in the SDP "a=fmtp" attribute by copying them directly from the MIME media type string as a semicolon separated with parameter=value, where parameter is "mode", and values can be 0, 20 or 30 (where 0 is reserved; 20 stands for preferred 20 ms frame size and 30 is reserved). An example of the media representation in SDP for describing iLBC when 20 ms frame size mode is used might be:

m=audio 49120 RTP/AVP 97
a=rtpmap:97 iLBC/8000
a=fmtp:97 mode=20
a=mptime:20

An example of the media representation in SDP for describing iLBC when 30 ms frame size mode is used might be:

m=audio 49150 RTP/AVP 99
a=rtpmap:99 iLBC/8000
a=mptime:30

As indicated in the example, when "mode" parameter in SDP "a=fmtp" attribute is not present, 30 ms frame size mode shall be applied.

7.6.2 BV16 Session Description

Parameters are mapped to SDP in a standard way. When conveying information by SDP, the encoding name shall be "BV16" (the same as the MIME subtype 0).

An example of the media representation in SDP for describing BV16 when 20 ms frame size mode is used might be:

m=audio 3456 RTP/AVP 97
a=rtpmap: 97 BV16/8000
a=mptime: 20

7.6.3 G.722 Session Description

Parameters are mapped to SDP in a standard way. When conveying information by SDP, the encoding name shall be "G722". G.722 has a static payload type of 9 as specified in [7].

Following is an example of the media representation in SDP for describing G.722 (using static payload type) when 20 ms frame size mode is used:

m=audio 3456 RTP/AVP 9
a=ptime: 20

Alternatively, the dynamic payload type may be used. In that case, the media representation would be:

m=audio 3456 RTP/AVP 99
a=rtpmap: 99 G722-64/8000
a=ptime: 20

8 Video Requirements

8.1 Overview

Packet-based video applications are one of the major potential enhancements to an IPCablecom service offering. Residential and business video conferencing, distance learning, and distance selling are just a few of the applications possible.

Yet this technology is nascent, and the precise content, form, and technology delivery for mass-market video applications is still gestating. The goal at this point for the IPCablecom effort is to clarify minimum video requirements for the most important current or anticipated interactive video applications, providing guideposts for implementations to maximize interoperability and customer satisfaction.

This clause addresses details of video communication over the IPCablecom network-in particular, the video codec requirements. The H.261 [i.9] and H.263 [i.10] Recommendations (as well as H.245 [i.8], or a functionally equivalent specification) are the basis and reference for this specification; highlights of these recommendations important to IPCablecom are illustrated here. Additionally, issues that have dependencies upon other IPCablecom resources, such as signalling and quality-of-service (QoS), are outlined.

8.2 IPCablecom Video Devices

The IPCablecom Multimedia Terminal Adapter 2 (MTA-2) offers video in addition to audio communication. The functional requirements of MTA-2 will be specified in the future.

8.3 Video Encoder Requirements

The video encoder provides a self-contained digital bitstream that may be combined with a media bitstream and/or signals. The video decoder performs the reverse process. Pictures are sampled at an integer multiple of the video-line rate. This sampling clock and the digital network clock are asynchronous. The transmission clock is provided externally. The video bitrate may be variable. In H.263 [i.10], no constraints on the video bitrate are given; the terminal or the network, as determined by the CMS or gatekeeper, provide constraints.

For reasons of interoperability, all IPCablecom MTA-2 terminals providing video communications shall be capable of encoding and decoding video according to H.261. This will permit video communication without the transcoding of video with terminals across the other networks, such as H.320 [i.11] terminals across an ISDN network or an H.324 [i.12] terminal across a PSTN network. The use of H.261 establishes a common denominator across all communication networks and retains backward compatibility with existing systems.

However, H.263 [i.10] is the preferred video codec and recommended for use in IPCablecom systems for a variety of reasons. Therefore, all IPCablecom MTA-2 terminals providing video communications shall also be capable of encoding and decoding video according to H.263. The most important improvement in H.263 is the advancement in motion estimation accuracy to a half-pixel, yielding a lower bit-per-picture requirement at a given bitrate. This, as well as several other advancements in the H.263 baseline codec and Annexes listed below, result in a higher frame rate and/or resolution at a given bitrate versus H.261.

8.4 Video Format Requirements

As stated in ITU-T Recommendation H.263 [i.10]:

"To permit a single recommendation to cover use in and between regions using 625- (PAL) and 525- (NTSC) line television standards, the source coder operates on pictures based on a common intermediate format (CIF). The standards of the input and output television signals, which may, for example, be composite or component, analogue or digital and the methods of performing any necessary conversion to and from the source coding format are not subject to recommendation."

The possible resolutions for the H.261 are CIF and quarter common intermediate format (QCIF). The possible resolutions for H.263 are sub-QCIF (SQCIF), QCIF, CIF, 4CIF, and 16CIF. CIF and QCIF are defined in H.261; SQCIF, 4CIF and 16CIF are defined in H.263.

Table 5: Number of Pixels Per Line and Number of Lines for Each Picture Format

Picture Format	Number of pixels for luminance (dx)	Number of lines for luminance (dy)	Number of pixels for chrominance (dx/2)	Number of lines for chrominance (dy/2)
SQCIF	128	96	64	48
QCIF	176	144	88	72
CIF	352	288	176	144
4CIF	704	576	352	288
16CIF	1 408	1 152	704	576

An MTA-2 shall support CIF and QCIF at a minimum. CIF is required for casual videoconferencing usage and is efficient for conferencing with a reasonable amount of motion at bitrates ranging from 128 kbps to 768 kbps. QCIF is required for interoperability with other endpoints not capable of encoding or decoding CIF, or if the MTA-2 is required to encode or decode two or more video streams in the case of a multi-point call.

MTA-2 implementations may employ SQCIF, 4CIF and 16CIF.

SQCIF is any active picture size less than QCIF, filled out by a black border, and coded in the QCIF format. SQCIF could be used for multiple encode or decode streams, as well as interoperability with a very low bit rate channel such as wireless.

4CIF and 16CIF are suitable for applications requiring very high resolution per frame as 4CIF exceeds the resolution of NTSC displays and 16CIF is four times this format. Examples of applications for 4CIF and 16CIF are high-resolution snapshots, document cameras, corporate business conferencing, and broadcast-quality streaming video. Snapshots and still frames at these resolutions are possible at all frame rates. Motion video at these resolutions typically will require a very high bit rate depending upon the desired frame rate.

For all these formats, the pixel aspect ratio is the same as that of the CIF format.

NOTE: The resulting picture aspect ratio for H.263 SQCIF is different from the other formats.

Other video codecs, and other picture formats, may also be employed, depending upon mutual device negotiation. The MTA-2 terminal optionally may send more than one video channel at the same time, for example, to convey the speaker and a second video source. The MTA-2 terminal optionally may receive more than one video channel at the same time, for example, to display multiple participants in a distributed multipoint conference.

The video bitrate, picture format, and algorithm options, which can be accepted by the decoder, shall be defined during the capability exchange. The encoder may transmit any or all options that are within the decoder capability set. The decoder should generate requests for preferred modes, but the encoder may ignore these requests if they are not mandatory modes. Decoders indicating capability for a particular algorithm option also shall be capable of accepting mandatory video bitstreams that do not make use of that option.

MTA-2 terminals shall be capable of operating in asymmetric video bit rates, frame rates, and picture resolutions (if more than one picture resolution is supported). For example, this will allow a CIF-capable terminal to transmit QCIF while receiving CIF pictures.

As stated in the H.263 recommendation, when each video logical channel is opened, the maximum operating mode to be used on that channel shall be signalled to the receiver. The maximum mode signalled includes maximum picture format, algorithm options, maximum codec bitrate, etc., as defined in H.263.

The header within the video logical channel indicates which mode, within the stated maximum, actually is used for each picture. For example, a video logical channel opened for CIF format may transmit CIF, QCIF, or SQCIF pictures, but not 4CIF or 16CIF. A video logical channel may negotiate subsets of options, but shall not use options that were not signalled.

8.5 H.263 Annexes

In addition to the H.263 baseline codec, there are several annexes that can improve the picture quality (with respect to frame rate, resolution, and bit-per-pixel coding efficiency). All of these annexes may be supported as optional codec features. Brief descriptions (from the H.263 recommendation) of each of the annexes follow. In order to guide vendor development and to encourage the highest common denominator of video quality possible employing the H.263 Recommendation, the descriptions include recommendations of the applicability and/or usefulness of the H.263 annexes to the IPCablecom video codec effort.

Annex D - Unrestricted Motion Vector Mode

Does two things:

1) allows motion vectors to point outside the picture boundaries; and
2) allows for longer motion vectors. Adds some complexity in the motion estimation process, but the longer vectors may be useful for larger picture sizes.

Recommendation: MTA-2s should employ this mode.

Annex E - Syntax-based Arithmetic Coding

Describes an alternate method of coding VLC codeword symbols. Adds considerable complexity with only marginal gain in compression performance. May also suffer in the error resiliency department.

Recommendation: MTA-2s should not employ this mode.

Annex F - Advanced Prediction Mode

Main contribution is overlapped block motion compensation (OBMC), which yields much smoother prediction. There is a considerable increase in complexity, and Annex J (below) accomplishes much the same thing (with lower complexity). Despite this, it is still beneficial or, at the very least, should be the first "high complexity option" chosen.

Recommendation: MTA-2s should employ this mode.

Annex G - PB-Frames Mode

Describes a method for increasing temporal resolution (especially for lower bitrates) through the use of bidirectionally predicted B-frames. Adds complexity and delay, plus the B-frames tend to take a hit in quality.

Recommendation: MTA-2s should not employ this mode.

Annex H - Forward Error Correction for Coded Video Signal

Describes a method for forward error correction (FEC) for the H.263 video signal.

Recommendation: MTA-2s should not employ this mode.

Annex I - Advanced INTRA Coding Mode

Describes an alternate method of coding INTRA blocks. Requires only a small increase in complexity, but yields only minimal quality gain.

Recommendation: MTA-2s should employ this mode.

Annex J - Deblocking Filter Mode

Describes a simple edge-deblocking filter used inside the video-coding loop (as opposed to a non-standardized postprocessing filter). Resulting quality is comparable in many cases to that obtained using Annex F (above), but with far fewer and much simpler calculations.

Recommendation: MTA-2s should employ this mode.

Annex K - Slice Structured Mode

Permits the use of (mostly) arbitrary resynchronization points within a picture (as opposed to GOB resynch points only), making it quite amenable to packet-based transports. Increases error resilience with little gain in complexity. Small (subpicture-duration) increase in delay, just as if GOB resync points had been used.

Recommendation: MTA-2s should employ this mode.

Annex L - Supplemental Enhancement Information

Describes the format for sending supplemental information related to a picture or pictures, e.g. picture freeze/release. A necessity for multipoint communications. Negligible increase in complexity.

Recommendation: MTA-2s should employ this mode.

Annex M - Improved PB-Frames Mode

Similar to Annex G (above), but with an improved methodology. Same general shortcomings (i.e. complexity, delay), however.

Recommendation: MTA-2s should not employ this mode.

Annex N - Reference Picture Selection Mode

Modifies the temporal prediction process by allowing the use of pictures other than the immediately preceding picture as a reference picture for prediction. May be useful in error-prone environments. Increases complexity and storage requirements. Requires a back channel.

Recommendation: MTA-2s may employ this mode.

Annex O - Temporal/SNR/Spatial Scalability

Describes methods to implement temporal (frame rate), SNR (picture quality), and/or spatial (picture size) scalability. In other words, being able to decode a sequence at multiple levels of perceived quality, i.e. layered video codecs. Substantial increase in complexity and bitrate, as well as an increase in delay in many cases.

Recommendation: MTA-2s should not employ this mode.

Annex P - Reference Picture Resampling

Describes a process in which the reference picture used for prediction is resampled ("warped") prior to prediction.

Recommendation: MTA-2s should not employ this mode.

Annex Q - Reduced Resolution Update Mode

Allows reduced (spatial) resolution updates to a reference picture having a higher resolution.

Recommendation: MTA-2s should not employ this mode.

Annex R - Independently Segmented Decoding Mode

Improves error resilience by localizing errors to only a segment (or slice; see Annex K, above) of a picture. Significantly improves error robustness in the presence of packet loss. Yields some loss in compression efficiency, however, as well as a moderate increase in complexity.

Recommendation: MTA-2s should employ this mode.

Annex S - Alternative INTER VLC Mode

Specifies an alternate VLC coding table for INTER-coded pictures in order to increase compression efficiency. Minimal improvement, at the expense of error detection capability (VLC table switching relies on the number of decoded coefficients being greater than 64, removing the ability to detect this sort of run-length error).

Recommendation: MTA-2s should not employ this mode.

Annex T- Modified Quantization Mode

Modifies the operation of the quantizer, e.g. step size, DCT coefficient range. Improves colour representation (especially in high-motion sequences) and adds additional error detection capability. Minimal increase in complexity.

Recommendation: MTA-2s should employ this mode.

A summary of these recommendations is presented in the table below. Also listed (for purposes of comparison only) are the three levels of preferred mode support described in Appendix II of H.263.

Table 6: H.263 Annexes and their Applicability to IPCablecom

Annex	H.263 Preferred Modes			IPCablecom?
	Level 1	Level 2	Level 3	
D		x	x	Y
E				N
F			x	Y
G				N
H				N
I	x	x	x	Y
J	x	x	x	Y
K		x	x	Y
L	x	x	x	Y
M			x	N
N				Y/N
O				N
P		x	x	N
Q				N
R			x	Y
S			x	N
T	x	x	x	Y

8.6 Multipoint Conferencing Support

In addition to the basic operation for encoding and decoding video streams, the MTA-2 may include support for multipoint conferences. If so, there are several commands particular to the video codec that enable multipoint support. These are:

8.6.1 Freeze Picture Request

Causes the decoder to freeze its displayed picture until a freeze picture release signal is received or a time-out period of at least six seconds has expired. The transmission of this signal is by external means.

8.6.2 Fast Update Request

Causes the encoder to encode its next picture in INTRA mode with coding parameters to avoid buffer overflow. The transmission method for this signal is by external means.

8.6.3 Freeze Picture Release

A signal from an encoder that has responded to a fast update request and allows a decoder to exit from its freeze picture mode and display decoded pictures in the normal manner. This signal is transmitted in the picture header of the first picture coded in response to the fast update request.

8.6.4 Continuous Presence Multipoint (CPM)

In H.263, a negotiable CPM mode is provided in which up to four independent H.263 QCIF bitstreams can be multiplexed as independent "sub-bitstreams" into one new video bitstream. Capability exchange for this mode is signalled by external means. Each sub-bitstream is considered as a normal H.263 bitstream and therefore shall comply with the capabilities that are exchanged by external means. The information in each individual bitstream is also completely independent from the information in the other bitstreams; for example, the picture rates for the different H.263 bitstreams may be different from one another.

8.7 Signalling Messages

At the time of this specification, the precise signalling protocol for all client devices has not been specified, but the following discussion demonstrates the necessary signals, whatever the protocol.

H.245 [i.8] provides an example of essential signalling components vital to an MTA-2 video call. Not only can H.245 be used for the exchange of capabilities at the initialization of a call, it may also be used during a call for several video and conference-centric commands. A list of mandatory (M) and optional (O) signals from the H.245 command set is shown below for receiving and transmitting MTA-2s. The mandatory commands (or their functional equivalents) shall be implemented in the IPCablecom signalling system.

Table 7: H.245 Commands that are Applicable to IPCablecom

Message	Receiving MTA Status	Transmitting MTA Status
Send Terminal Capability Set	M	M
Encryption	O	O
Flow Control	M	O
End Session	M	M
Miscellaneous Commands		
Equalize Delay	O	O
Zero Delay	O	O
Multipoint Mode Command	M	O
Cancel Multipoint Mode Command	M	O
Video Freeze Picture	M	O
Video Fast Update Picture	M	O
Video Fast Update GOB	M	O
Video Fast Update MB	M	O
Video Temporal Spatial Trade Off	O	O
Video Send Sync Every GOB	O	O
Video Send Sync Every GOB Cancel	O	O
MCLocationIndication	M	O
Terminal ID Request	O	O
Terminal List Request	O	O
Broadcast Me	O	O
Cancel Broadcast Me	O	O
Make Terminal Broadcaster	O	O
Send This Source	O	O
Cancel Send This Source	O	O
Drop Terminal	O	O
Make Me Chair	O	O
Cancel Make Me Chair	O	O
Drop Conference	O	O
Enter H.243 Password	O	O
Enter H.243 Terminal Id	O	O
Enter H.243 Conference ID	O	O
Request Terminal ID	O	O
Terminal ID Response	O	O
Terminal List Response	O	O
Video Command Reject	O	O
Make Me Chair Response	O	O
NOTE: M = mandatory. O = optional.		

9 RTCP Requirements and RTCP Usage

9.1 RTP Requirements

The voice and fax/modem pass-through media flows shall be transported using IETF Real-Time Transport Protocol (RTP) and Real-Time Transport Control Protocol (RTCP) as defined in RFC 1889 [16] and RFC 1890 [7]. All IPCablecom devices supporting RTP (e.g. MTAs, trunking gateways, audio servers) shall support RTCP as defined in RFC 1889 [16] and RFC 1890 [7] and profiled in this clause.

IPCablecom endpoints that perform mixing of RTP streams may transmit contributing source lists (CSRC). This requirement is intended to allow mixers to omit CSRC lists, in compliance with RFC 1889 [16] and RFC 1890 [7], to avoid resource management issues that may arise from contributing sources joining and leaving sessions, resulting in dynamic, variable-length RTP packet headers. These issues remain for further study.

IPCablecom endpoints shall accept RTP packets that contain contributing source lists (CSRC). This requirement is intended to allow endpoints to interoperate successfully with non-IPCablecom mixers and IPCablecom mixing endpoints that transmit CSRC lists.

9.2 RTCP Requirements

To facilitate vendor interoperability, the following RTCP profile has been defined for IPCablecom-compliant endpoints. In the event that a discrepancy arises between the RFCs and this profile, this profile will take precedence.

9.2.1 General Requirements of the IPCablecom RTCP Profile

IPCablecom endpoints shall send RTCP messages, as described in RFC 1889 [16] and RFC 1890 [7] and profiled below.

Endpoints may start transmitting RTCP messages as soon as the RTP session has been established, even if RTP packets are not being sent or received. An RTP session is considered established once each endpoint has received a remote connection descriptor. Furthermore, an IPCablecom endpoint shall start transmitting RTCP messages if it receives an RTCP message. Once started, the endpoint shall not stop sending RTCP messages, except for the cases identified below.

To avoid unnecessary network traffic, endpoints may stop sending RTCP packets to a remote endpoint if an ICMP port unreachable or another ICMP destination unreachable error (i.e. ICMP error type 3) is returned from the network for that RTCP destination.

To avoid unnecessary network traffic, endpoints may stop sending RTCP packets to a remote endpoint if no RTCP packets have been received within five (5) report transmission intervals. This requirement allows the endpoint to stop sending RTCP packets to endpoints that simply receive and discard RTCP reports.

An RTCP transmission interval calculation procedure is outlined in clause 9.2.

IPCablecom endpoints shall receive RTCP messages, if sent by the remote communication peers. IPCablecom endpoints shall not require them. That is, call state in general and RTP flows in particular shall not be affected by the absence of one or more RTCP messages. This requirement is intended to facilitate interoperability with non-IPCablecom endpoints.

By default, RTCP messages receive best effort treatment on the network. RTCP messages may receive better than best-effort treatment on the network. QoS-enhanced treatment is possible, but is not required by this profile. RTCP packets that are transmitted with best effort treatment may be delayed or lost in the network. As such, any application that attempts to use RTCP for accurate estimate of delay and latency, or to provide liveliness indication, for example, needs to be tolerant of delay or packet loss. If delay or packet loss cannot be tolerated, the application can use QoS enhanced treatment for RTCP, but this requires establishment of additional service flow(s), probably separate from the service flows established to carry the RTP stream. Setting up additional flows has significant implications for HFC access network bandwidth utilization, admission control, call signalling, and DOCSIS signalling, and remains for further study.

SSRC (Synchronization Source) collision detection and resolution is OPTIONAL for IPCablecom endpoints that are capable of unambiguously distinguishing between media packets and reports that they send and those that it receives. If an endpoint can handle SSRC collisions without affecting the integrity of the session, the endpoint may ignore SSRC collisions. In particular, SSRC collision detection and resolution is OPTIONAL for endpoints that are establishing unicast, point-to-point connections carrying one RTP stream, as is the case in current IPCablecom connections. If SSRC collision detection and resolution is supported, one or both of the endpoints shall resolve SSRC collisions as follows:

1) send BYE;

2) select new SSRC;

3) send Sender Description with new SSRC.

SSRC collision detection and resolution is OPTIONAL for IPCablecom endpoints that perform mixing for multiple remote endpoints when CSRC lists are not transmitted in the mixed packets. When CSRC lists are transmitted, the mixing endpoint shall detect and resolve SSRC collisions.

Future IPCablecom connections may involve multiple, simultaneous RTP streams, and require resolution of SSRC collisions. In this case responsibility for this resolution falls to the two colliding senders. One or both of these parties shall resolve SSRC collisions as follows:

1) send BYE;

2) select new SSRC;

3) send Sender Description with new SSRC.

The following defines normative requirements placed on specific RTCP protocol messages:

SDES (Source Description): CNAME objects shall not contain identity information (see definition below); CNAME field shall be a cryptographically-random value generated by the endpoint in such a manner that endpoint identity is not compromised and shall change on a per-session basis; NAME, EMAIL, PHONE, LOC objects should not be sent and, if sent, shall not contain identity information. This requirement is intended to satisfy the requirements of [16] with respect to the CNAME field, and at the same time satisfy legal and regulatory requirements for maintaining subscriber privacy, for example, when caller id blocking must be performed. This requirement is imposed because not all RTCP messages may be encrypted, as described in the IPCablecom Security Specification [i.1].

SR (Sender Report): shall be sent by IPCablecom endpoints transmitting RTP packets (as described in [16]), except as previously described when errors occur or the remote endpoint does not send RTCP packets, in which case they may be sent.

RR (Receiver Report): shall be sent with report blocks if receiving but not sending RTP packets (as described in [16]) and shall be sent without report blocks if not sending or receiving RTP packets, except as previously described when errors occur or the remote endpoint does not send RTCP packets, in which case they may be sent.

APP (Application-Defined): may be sent as implementation needs dictate and shall not contain identity info. Endpoints shall ignore and silently discard APP messages with unrecognized contents.

BYE (Goodbye): shall be sent upon RTP connection deletion or when renegotiating SSRC upon collision detection and resolution (see below). Endpoints shall send BYE commands when the application needs to discontinue use of an SSRC and start a new SSRC, for example, on codec change.

NOTE 1: Codec change is an example only, since in some implementations, the endpoint may not need to change SSRC when changing codec.)

Endpoints shall not use BYE messages to indicate or detect any call progress condition. For example, endpoints shall not tear down RTP flows based on BYE, but shall update RTCP/RTP state as per RFC 1889 [16]. This requirement is intended to ensure that all call progress conditions, such as on-hook notifications, are signalled using the higher-level IPCablecom signalling protocol, such as Network-based Call Signalling (NCS).

NOTE 2: Identity information refers to any token (e.g. name, e-mail address, IP address, phone number) which may be used to reveal the particular subscriber or endpoint device in use.

9.2.2 Security Requirements for RTP and RTCP in IPCablecom

IPCablecom endpoints shall not conform to the security requirements described in the RTP/RTCP RFC and drafts. Instead, IPCablecom endpoints shall implement RTP and RTCP security as specified in the IPCablecom Security Specification [i.1].

9.2.3 Extended RTCP Reports

The RTCP XR VoIP Metrics Report Block as defined in [29] shall be sent by endpoints if negotiated on a given connection as defined in IPCablecom Network Based Call Signalling Protocol Specification [3] and Trucking Gateway Control Protocol Specification [21]. IPCablecom endpoints may send other RTCP XR payload types. IPCablecom endpoints that are capable of sending RTCP XR reports shall be capable of receiving, interpreting and parsing RTCP XR VoIP Metrics reports.

9.2.3.1 Reporting Call Quality Metrics using RTCP XR

9.2.3.1.1 RTCP XR VoIP Metrics Requirements

The RTCP XR VoIP Metrics [29] report provides a set of performance metrics that can be helpful in diagnosing problems affecting call quality. RTCP XR is a media path reporting protocol, i.e. messages are exchanged between endpoints, however they may be captured by intermediate network probes or analyzers, or potentially by embedded monitoring functionality in CMTS and routers. The RTCP XR VoIP metrics are also reported when the connection is deleted.

IPCablecom endpoints shall exchange RTCP XR VoIP Metrics reports during active RTP sessions if negotiated and shall concatenate RTCP XR payloads with RTCP SR and RR payloads, following rules for transmission intervals [16].

IPCablecom endpoints that support the RTCP XR VoIP Metrics payload shall measure or compute the reported values of the metrics as defined in clauses 9.2.3.1.2 to 9.2.3.1.6 of the present document.

9.2.3.1.2 Definition of Metrics related to Packet Loss and Discard

The VoIP Metrics [29] payload contains six metrics related to packet or frame loss and discard. An average packet loss rate and an average packet discard rate report the proportion of packets lost or discarded on the call to date. A set of four burst parameters report the distribution of lost and discarded packets occurring during burst periods and gap periods.

RTCP XR views a call as being divided into bursts, which are periods during which the combined packet loss and discard rate is high enough to cause noticeable call quality degradation (generally over 5 percent loss/discard rate), and gaps, which are periods during which lost or discarded packets are infrequent and hence call quality is generally acceptable. A parameter Gmin is associated with these definitions and shall be set to 16 within IPCablecom systems.

Table 8: Metrics Related to Packet Loss and Discard

METRIC	Description	Range
Loss Rate	Proportion of packets lost within the network	0 to 0,996
Discard Rate	Proportion of packets discarded due to late arrival	0 to 0,996
Burst Loss Density	Proportion of packets lost and discarded during burst periods	0 to 0,996
Gap Loss Density	Proportion of packets lost and discarded during gap periods	0 to 0,996
Burst Duration	Average length of burst periods (ms)	0 to 65,535
Gap Duration	Average length of gap periods (ms)	0 to 65,535
Gmin	Parameter used to define burst periods	16

An IPCablecom endpoint when using RTCP XR shall provide these parameters as defined in table 8.

9.2.3.1.3 Definition of Metrics Related to Delay

The VoIP Metrics payload includes two delay metrics [29]. The Round Trip Delay is the delay between RTP interfaces, as typically measured using RTCP Sender Report (SR) or Receiver Report (RR) [16]. The End System Delay incorporates the vocoder encoding and decoding delay, the packetization delay, and the current nominal delay due to the jitter buffer.

Table 9: Metrics related to Delay

Metric	Description	Range
Round Trip Delay	Packet path round trip delay (mS)	0 to 65,535
End System Delay	Round trip delay within end system (mS)	0 to 65,535

An IPCablecom endpoint when using RTCP XR shall provide the parameters as defined in table 9.

NOTE: This requires an SR or RR exchange prior to the inclusion of an XR payload into an RTCP message.

9.2.3.1.4 Definition of Metrics related to Signal

The Signal Level, Noise Level and estimated Residual Echo Return Loss are intended to support the diagnosis of problems related to loss plan or PSTN echo. The intent is to report useful information that would typically be available from a vocoder or echo canceller rather than to impose the overhead of additional measurement algorithms on cost sensitive endpoints.

The signal and noise level estimates are expressed in dBm0 with reference to a digital milliwatt and relate to the received VoIP packet stream. The effects of a low or high signal level or a high noise level will affect the user at the endpoint reporting this metric.

The Residual Echo Return Loss is the echo canceller's estimate of the line echo remaining after the effects of echo cancellation, echo suppression and non-linear processing; note that this will in general not represent an accurate measurement of the residual echo but can provide a useful indication of the presence of echo problems. Echo occurring on the endpoint reporting this metric will be heard by the user at the remote endpoint, if significant delay is present on the call.

Table 10: Metrics due to Signal

METRIC	Description	Range
Signal Level	RMS Signal level during active speech periods (dBm0) As defined in [30] and [31].	-30 to +3
Noise Level	RMS Noise level during silence periods (dBm0) As defined in [30] and [31].	-40 to -70
Residual Echo Return Loss	Estimated Echo Return Loss (after effects of echo canceller and NLP) from the local line echo canceller (dB) As defined in ITU-T Recommendation G.168 [19].	0 to 80

An IPCablecom endpoint when using RTCP XR shall provide Signal Level and Noise as defined in table 10.

An IPCablecom endpoint equipped with an echo canceller and when using RTCP XR shall provide the Residual Echo Return Loss metric as defined in table 10.

9.2.3.1.5 Definition of Metrics related to Call Quality

Call quality metrics are useful when assessing the overall quality of a call [29]. A listening quality metric represents the effects of vocoder distortion, lost and discarded packets, noise and signal level on user perceived quality. A conversational quality metric also includes the effects of delay and echo on user perceived quality. Call quality metrics are often expressed in terms of a transmission quality rating or R factor (from the E Model [32]) or in terms of Mean Opinion Score (MOS).

The maximum range of an R factor is 0 to 100 for narrowband voice transmission.

NOTE 1: However, for wideband transmission the upper range can be greater than 100.

The R factor defined in the ITU E Model is a conversational quality metric however it can be used to estimate conversational and listening quality MOS scores. The basic equation for determining an R Factor is:

$$R = Ro - Is - Id - Ie,eff + A$$

Ro reflects the effects of noise and loudness, Is the effects of impairments occurring simultaneously with speech, Id the effects of delay related impairments and echo, Ie,eff the "equipment impairment" factors and A is used to correct for the convenience of services such as cellular networks.

Strictly, a MOS can only be obtained from subjective testing, however the MOS scale represents a convenient and well understood scale, and hence is often used. ITU-T Recommendation G.107 [32] defines an equation for converting an R factor into a MOS score.

NOTE 2: That this produces MOS scores slightly higher than those typically reported from subjective tests.

Table 11: Metrics related to Call Quality

Metric	Description	Range
R Factor	Conversational Transmission Quality Rating	0 to 100
External R Factor	R factor for an attached external network	0 to 100
MOS-LQ	Estimated listening quality MOS (x10)	10 to 50
MOS-CQ	Estimated conversational quality MOS (x10)	10 to 50

An IPCablecom endpoint when using RTCP-XR shall provide the R Factor, MOS-LQ and MOS-CQ metrics and may provide an External R Factor.

An IPCablecom endpoint when using RTCP XR shall calculate R Factors using G.107 at a minimum [32].

An IPCablecom endpoint when using RTCP XR shall calculate the Ro, Is and Id parameters based on the Signal Level, Noise Level, Round Trip Delay and End System Delay values determined locally and the Residual Echo Return Loss, End System Delay and Signal Level reported by the remote endpoint.

In order to determine Ro, Is and Id the following mappings of measured parameters shall be used.

E Model No parameter = Noise Level.

E Model SLR parameter = SLR(Remote) = -15 - Signal Level (Local).

SLR(Local) = -15 - Signal Level (Remote).

The Signal Level (Remote) is obtained from a received RTCP XR message from the remote endpoint. If no RTCP XR message has been received then E Model default value for SLR MUST be assumed. For more information refer to [32].

E Model TELR parameter = SLR(Local) + RERL(Remote) + RLR(Local)

The RERL (Remote) is obtained from a received RTCP XR message from the remote endpoint. If no RTCP XR message has been received then E Model default value for TELR MUST be assumed. For more information refer to [32].

Total Delay = End System Delay(Remote) + Round Trip Delay + End System Delay(Local)

The End System Delay (Remote) is obtained from a received RTCP XR message from the remote endpoint. If no RTCP XR message has been received then the remote end system delay shall be assumed to be equal to the local end system delay. For more information refer to [32].

Also the following equations below explain how to take measurements above and apply those to the E-model input parameters. For more information refer to [32].

E Model Ta = T = Total Delay / 2.

E Model Tr = Total Delay.

E Model Ppl = Average packet loss and discard rate for call.

Other E Model parameters should be set to defaults or to predetermined values for the endpoint. For more information refer to [32].

An IPCablecom endpoint when using RTCP XR shall calculate the Ie,eff parameter using the function defined in G.107 [32]. However, the IPCablecom endpoint shall use the Ie and Bpl parameters defined in table 12 for the Vocoder and PLC combinations listed.

Table 12: Ie and Bpl parameters for IPCablecom Vocoders

Vocoder	Bit rate	PLC	Ideal R	Ideal MOS	Ie	Bpl
G.711 A/U	64k	Appendix 1 [4]	93	4,4	0	34
G.728 10 ms	16k	Per G.728 Annex I [22]	89	4,1	7	17
G.728 20 ms	16k	Per G.728 Annex I [22]	89	4,1	7	15
G.729 Annex E 10 ms	11,8k	Per G.729 [23]	88	4,1	4	20
G.729 Annex E 20 ms	11,8k	Per G.729 [23]	88	4,1	4	19
ILBC 20 ms	15,2k	Per [36]	80	3,9	10	34
ILBC 30 ms	13,3k	Per [36]	78	3,8	12	27
BV16 10 ms	16k	Per [38]	88	4,2	5	25
BV16 20 ms	16k	Per [38]	88	4,2	5	23

An IPCablecom endpoint when using RTCP XR shall calculate MOS-LQ using the R to MOS mapping function defined in G.107 [32] applied to the value (R - Id).

An IPCablecom endpoint when using RTCP XR shall calculate MOS-CQ using the R to MOS mapping function defined in G.107 [32] applied to the value R.

Ie and Bpl values for new Codecs can be determined using objective and subjective test data. An example procedure for determining these values is given below:

a) Use ITU-T Recommendation P.862 [33] to build a table of objective test score vs. packet loss rate for a range of at least 0 to 10 percent loss. For each packet loss rate use at least eight source audio files, encode each file using the codec under test, apply the packet loss rate and then decode the file using the codec under test with the associated packet loss concealment algorithm. Use P.862 to compare the impaired output files with the source files and average the results for each packet loss rate.

b) Determine the Ie value using the objective test scores for 0 percent loss. This may be obtained by iteratively searching for the Ie value that, when converted to an R factor and then an estimated P.862 score, gives the closest match to the measured P.862 score. Alternatively, the Ie value may be obtained by comparing the P.862 [33] score with other codecs with known Ie factor.

$$R_{adj} = R + (94 - R) / 3 - 3 - 115 / (15 + ABS (85 - R)) + 40 / (95 - R)^2$$

$$\text{Estimated PESQ score} = 1 + 0{,}033R_{adj} + R_{adj}(100-R_{adj})(R_{adj}-60) \times 0{,}000007$$

c) Determine the Bpl value using the objective test scores for other packet loss rates. This may be obtained by iteratively searching for the Bpl value that, when converted to an R factor and then an estimated P.862 [33] score, gives the closest match to the measured P.862 [33] score. Alternatively, the Bpl value may be obtained by comparing the P.862 [33] score curve with other codecs with known Bpl factor.

d) It is generally advisable to compare the curve of estimated MOS score (derived per G.107 [32]) with available ACR test data (if available) in order to verify values.

9.2.3.1.6 Definition of Parameters related to endpoint configuration

These parameters in table 13 describe some key configuration parameters of the IPCablecom endpoint, that are useful in monitoring service quality and identifying some types of configuration related problems.

Table 13: Parameters related to endpoint configuration

METRIC	Description	Range
PLC Type	Type of packet loss concealment algorithm:	UnspecifiedDisabledEnhancedStandard
Jitter Buffer Type	Type of jitter buffer (fixed or adaptive)	UnknownReservedNon-adaptiveAdaptive
Jitter Buffer Rate	Rate of adjustment of an adaptive jitter buffer	0 to 15
Jitter Buffer- Nominal Delay	Nominal delay applied to received packets by the jitter buffer for packets arriving on time	0 to 65,535
Jitter Buffer - Maximum Delay	Maximum delay applied to received packets by the jitter buffer	0 to 65,535
Jitter Buffer - Absolute Max Delay	Maximum delay size that an adaptive jitter buffer can reach	0 to 65,535

An IPCablecom endpoint when using RTCP XR shall provide values to all Parameters as defined in table 13.

Annex A (informative): Codec Comparison Tables

The following three tables summarize standard speech coder characteristics.

Some of the data in the three tables are obtained from Current Methods of Speech Coding [6].

Table A.1: ITU IETF and SCTE Speech Coders

Standards Body	ITU	ITU	ITU	ITU	ITU	ITU	ITU	ITU	ITU	IETF1/	SCTE
Recom-mendation	G.711	G.726	G.728	G.729	G.729A	G.729D	G.729E	G.723.1	G.722	iLBC	BV16
Coder Type	Companded PCM	ADPCM	LD-CELP	CS-ACELP	CS-ACELP	CS-ACELP	CS-ACELP	MPC-MLQ & ACELP	SB-ADPCM	FB-LPC	TSNFC
Dates	1972	1990	1992/4	1995	1996	1998	1998	1995	1988	2002	2003
Bitrate	64 kb/s	16 kb/s to 40 kb/s	16 kb/s	8 kb/s	8 kb/s	6,4 kb/s	11,8 kb/s	6,3 kb/s & 5,3 kb/s	48 kb/s, 56 kb/s, 64 kb/s	15,2 kb/s & 13,3 kb/s	16 kb/s
Peak Quality (see note 2)	Toll	≤Toll	Toll	Toll	Toll	< Toll	Toll	≤Toll	>Toll	Toll	Toll
Background Noise (see note 3)	Toll	≤ Toll	Toll	≤ Toll	≤ Toll	< Toll	Toll	≤ Toll	N/A	Toll	Toll
Tandem (see note 4)	Toll	Toll	Toll	< Toll	< Toll	< Toll	Toll	< Toll	N/A	< Toll	Toll
Frame Erasure (see note 5)	No mechanism	No mechanism	3 %	3 %	3 %	3 %	3 %	3 %	No mechanism	7 % and 5 %	5 %
Complexity (MIPS) (see note 6)	~0,35	~12	~36	~22	~13	~20	~27	~19	~10	~15 and ~18	~12
RAM (kword) (see note 7)	~0,01	~0,15	~2,20	~2,6	~2,6	~2,6	~2,6	~2,1	~1	~4	~2
Frame Size	0,125 ms	0,125 ms	0,625 ms	10 ms	10 ms	10 ms	10 ms	30 ms	0,0625	20 ms and 30 ms	5 ms
Look Ahead	0	0	0	5 ms	5 ms	5 ms	5 ms	7,5 ms	0	5 ms and 10 ms	0
Codec Delay (see note 8)	0,25 ms	0,25 ms	1,25 ms	25 ms	25 ms	25 ms	25 ms	67,5 ms	1,5625 ms	45 ms and 70 ms	10 ms

NOTE 1: The actual codec description is in the experimental standards track of IETF.
NOTE 2: Peak quality means clean input speech and clear channel for single encoding.
NOTE 3: Background noise refers to overall performance in background noises such as car noise, babble, office, and music.
NOTE 4: Tandems refer to the performance of the coder for multiple asynchronous encodings. Toll quality is defined as the performance of 32 kb/s G.726. Coders such as G.729, G.723.1, and others, are known to degrade more quickly with multiple tandems than G.726.
NOTE 5: Frame erasures refers to the rate at which the MOS score is approximately 0.5 MOS worse than the peak quality for that coder.
NOTE 6: Complexity is reported as MIPS (Million Instructions Per Second) and stated computational complexity numbers include one encoder and one decoder for the TI TMS320C54x architecture.
NOTE 7: RAM usage is reported in 16-bit words, the most common unit for fixed-point DSP implementations (due to 16-bit word length of many common DSPs). Stated RAM usage numbers include: "state memory RAM usage" of the encoder, the "state memory RAM usage" of the decoder and the worst case "temporary RAM usage" of the encoder and the decoder for the TI TMS320C54x architecture.
NOTE 8: Codec delay is equal to the sum of the look-ahead plus two times the frame size. The ITU uses this formula because it is assumed that the processing of a single device to encode and decode has to be accomplished in one frame-size time or less. The transmission time is a function of the network, as are other delays for a telephone call.

Table A.2: North American Wireless Speech Coders

Standards Body	TIA	TIA	TIA	TIA	TIA	ETSI	ETSI	ETSI
Recommendation	IS-54	IS-641	IS-96	IS-127	IS-733	GSM-(FR)	GSM-(HR)	GSM-(EFR)
System	TDMA	TDMA	CDMA	CDMA	CDMA	GSM	GSM	GSM
Coder Type	VSELP	ACELP	QCELP	ACELP	CELP	RPE-LTP	VSELP	ACELP
Dates	1990	1995	1993	1997	1997	1987	1994	1995
Bitrate	7,95 kb/s	7,4 kb/s	0,8 kb/s to 8,0 kb/s	0,8 kb/s to 8,55 kb/s	0,8 kb/s to 13,2 kb/s	13 kb/s	5,6 kb/s	12,2 kb/s
Peak Quality (see note 1)	= GSM-(FR)	Toll	= GSM-(FR)	Toll	Toll	<Toll	=GSM-(FR)	Toll
Background Noise (see note 2)	<< Toll	< Toll	<< Toll	< Toll	Toll	<Toll	< GSM-(FR)	Toll
Tandem (see note 3)	<< Toll	< Toll	<< Toll	< Toll	Toll	<< Toll	< GSM-(FR)	Toll
Frame Erasures (see note 4)	3 %	3 %	3 %	3 %	3 %	3 %	3 %	3 %
Complexity (MIPS) (see note 5)	~12	~15	~18	~25	~22	~5	~24	~18
RAM (kword) (see note 6)	~1,5	~2,5	~2	~2,5	~2,5	~1	~4	~4,6
Frame Size	20 ms	20 ms	20 ms	20 ms	20 ms	20 ms	20 ms	20 ms
Look Ahead	5 ms	5 ms	5 ms	5 ms	5 ms	0	4,4 ms	0
Codec Delay (see note 7)	45 ms	45 ms	45 ms	45 ms	45 ms	40 ms	44,4 ms	40 ms

NOTE 1: Peak quality means clean input speech and clear channel for single encoding.

NOTE 2: Background noise refers to overall performance in background noises such as car noise, babble, office, and music.

NOTE 3: Tandems refer to the performance of the coder for multiple asynchronous encodings. Toll quality is defined as the performance of 32 kb/s G.726. Coders such as G.729, G.723.1, and others, are known to degrade more quickly with multiple tandems than G.726.

NOTE 4: Frame erasures refers to the rate at which the MOS score is approximately 0,5 MOS worse than the peak quality for that coder.

NOTE 5: Complexity is reported as MIPS (Million Instructions Per Second) and stated computational complexity numbers include one encoder and one decoder for the TI TMS320C54x architecture.

NOTE 6: RAM usage is reported in 16-bit words, the most common unit for fixed-point DSP implementations (due to 16-bit word length of many common DSPs). Stated RAM usage numbers include: "state memory RAM usage" of the encoder, the "state memory RAM usage" of the decoder and the worst case "temporary RAM usage" of the encoder and the decoder for the TI TMS320C54x architecture.

NOTE 7: Codec delay is equal to the sum of the look-ahead plus two times the frame size. The ITU uses this formula because it is assumed that the processing of a single device to encode and decode has to be accomplished in one frame-size time or less. The transmission time is a function of the network, as are other delays for a telephone call.

G.729 was finalized in 1995 originally by the ITU to be a toll quality 8 kb/s standard. In that year, the ITU was requested to create a low-complexity coder for simultaneous voice and data. G.729A was created as a low-complexity version that is fully interoperable with G.729. G.729B is a speech/silence detector and comfort noise generator. It can be used with either G.729 or G.729A to provide an option for variable rate usage, also known as discontinuous transmission. G.729C contains the floating-point versions of G.729 and G.729A. G.729D is a 6,4 kb/s version of G.729. It was created to provide an optional lower rate that can be used briefly for periods of network congestion, or when more bits are needed for channel error protection. Its quality is less than that of G.729 or G.729A. G.729E is a higher rate version of G.729 designed to provide higher quality for background noise conditions, music, and tandems. It is a hybrid coder. It codes each frame two different ways and selects the method that appears to give the greater fidelity. Its forward-adaptive mode uses CS-ACELP. Its backward-adaptive mode features a 30^{th}-order backward-adaptive LPC synthesis filter and no pitch predictor. This mode is better for music, and it has greater complexity than the original G.729 coders.

Table A.3 is intended to provide essential access network bandwidth-related information for each codec listed. Although some of the listed codecs (e.g. G.711, G.726) are sample-based rather than frame-based, for anticipated purposes of flow management, frame-oriented packet sizes are listed. The three most important packet sizes are shown, corresponding to low latency (10, 20, and 30 ms) samples. Packet header overhead is calculated at 40 bytes, with 12 bytes RTP, 8 bytes UDP, and 20 bytes IP contributions.

NOTE: G.729E is shown at a byte-boundary 12 kb/s, which includes the 2 bits/frame not currently defined. Variable bit rate VAD implementations for each codec are not listed.

Table A.3: Bandwidth Attributes of Codecs

Codec	Bitrate (kb/s)	Byte/10 ms	Frm/Pkt	Byte/Pkt	Pkt/s	Byte/s	kb/s
G.711 - 10 ms	64	80	1	120	100	12 000	96
G.711 - 20 ms	64	80	2	200	50	10 000	80
G.711 - 30 ms	64	80	3	280	33,3	9 334	75
G.726.16 - 10 ms	16	20	1	60	100	6 000	48
G.726.16 - 20 ms	16	20	2	80	50	4 000	32
G.726.16 - 30 ms	16	20	3	100	33,3	3 334	27
G.726.24 - 10 ms	24	30	1	70	100	7 000	56
G.726.24 - 20 ms	24	30	2	100	50	5 000	40
G.726.24 - 30 ms	24	30	3	130	33,3	4 334	35
G.726.32 - 10 ms	32	40	1	80	100	8 000	64
G.726.32 - 20 ms	32	40	2	120	50	6 000	48
G.726.32 - 30 ms	32	40	3	160	33,3	5 334	43
G.726.40 - 10 ms	40	50	1	90	100	9 000	72
G.726.40 - 20 ms	40	50	2	140	50	7 000	56
G.726.40 - 30 ms	40	50	3	190	33,3	6 334	51
G.728 - 10 ms	16	20	1	60	100	6 000	48
G.728 - 20 ms	16	20	2	80	50	4 000	32
G.728 - 30 ms	16	20	3	100	33,3	3 334	27
G.729A - 10 ms	8	10	1	50	100	5 000	40
G.729A - 20 ms	8	10	2	60	50	3 000	24
G.729A - 30 ms	8	10	3	70	33,3	2 334	19
G.729E - 10 ms	12	15	1	55	100	5 500	44
G.729E - 20 ms	12	15	2	70	50	3 500	28
G.729E - 30 ms	12	15	3	85	33,3	2 834	23
iLBC - 20 ms	15,2	19	1	78	50	3 900	31
iLBC - 30 ms	13,3	16,67	1	90	33,3	3 000	24
BV16 - 10 ms	16	20	2	60	100	6 000	48
BV16 - 20 ms	16	20	4	80	50	4 000	32
BV16 - 30 ms	16	20	6	100	33,3	3 334	26,7
G.722 - 48 Kbps - 10 ms	48	60	1	100	100	10 000	80
G.722 - 48 Kbps - 20 ms	48	60	2	160	50	8 000	64
G.722 - 48 Kbps - 30 ms	48	60	3	220	33,3	7 334	58,7
G.722 - 56 Kbps - 10 ms	56	70	1	110	100	11 000	88
G.722 - 56 Kbps - 20 ms	56	70	2	180	50	9 000	72
G.722 - 56 Kbps - 30 ms	56	70	3	250	33,3	8 334	66,6
G.722 - 64 Kbps - 10 ms	64	80	1	120	100	12 000	96
G.722 - 64 Kbps - 20 m	64	80	2	200	50	10 000	80
G.722 - 64 Kbps - 30 m	64	80	3	280	33,3	9 334	74,6

Annex B (informative): Bibliography

IETF RFC 0821: "Simple Mail Transfer Protocol", August 1982.

IETF RFC 2104: "HMAC: Keyed-Hashing for Message Authentication", February 1997.

IETF RFC 1945: "Hypertext Transfer Protocol -- HTTP/1.0", May 1996.

IETF RFC 2068: "Hypertext Transfer Protocol -- HTTP/1.1", January 1997.

IETF RFC 3435: "Media Gateway Control Protocol (MGCP) Version 1.0", January 2003.

IETF RFC 2138: "Remote Authentication Dial In User Service (RADIUS)", April 1997.

IETF RFC 2139: "RADIUS Accounting", April 1997.

SIA DC-05-1999.09: "Digital Communication Standard - Ademco ® Contact ID Protocol - for Alarm System Communications", Security Industry Association Copyright 1999 - Ademco Group.

ETSI ES 201 235-2 (V1.2.1): "Access and Terminals (AT); Specification of Dual-Tone Multi-Frequency (DTMF) Transmitters and Receivers; Part 2: Transmitters".

ETSI ES 201 235-3 (V1.2.1): "Access and Terminals (AT); Specification of Dual-Tone Multi-Frequency (DTMF) Transmitters and Receivers; Part 3: Receivers".

PKT-SP-EM1.5-I03-070412, PacketCable 1.5 Event Messaging Specification, April 12, 2007, Cable Television Laboratories, Inc.

PKT-SP-PROV1.5-I04-090624, PacketCable 1.5 MTA Device Provisioning Specification, June 24, 2009, Cable Television Laboratories, Inc.

PKT-SP-TGCP1.5-I03-070412, PacketCable 1.5 PSTN Gateway Call Signaling Protocol Specification, April 12, 2007, Cable Television Laboratories, Inc.

PKT-SP-ASP1.5-I02-070412, PacketCable 1.5 Audio Server Protocol Specification, April 12, 2007, Cable Television Laboratories, Inc.

ITU-T Recommendation G.729 Annex A: "Reduced Complexity 8 kbit/s CS-ACELP Speech Codec", November 1996.

ITU-T Recommendation H.243 (February 2000): "Procedures for establishing communication between three or more audiovisual terminals using digital channels up to 1920 kbit/s".

History

Document history		
V1.1.1	April 2011	Publication