International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

J.161
(06/2007)

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

IPCablecom

Audio and video codec requirements and usage for the provision of bidirectional audio services over cable television networks using cable modems

ITU-T Recommendation J.161

ITU-T

International Telecommunication Union

ITU-T Recommendation J.161

Audio and video codec requirements and usage for the provision of bidirectional audio services over cable television networks using cable modems

Summary

ITU-T Recommendation J.161 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. This Recommendation also specifies the performance required in client devices to support future IPCablecom codecs. Additionally, this Recommendation describes a suggested methodology for optimal network support for codecs.

Source

ITU-T Recommendation J.161 was approved on 20 June 2007 by ITU-T Study Group 9 (2005-2008) under the ITU-T Recommendation A.8 procedure.

FOREWORD

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

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

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

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure e.g. interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had received notice of intellectual property, protected by patents, which may be required to implement this Recommendation. However, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database at http://www.itu.int/ITU-T/ipr/.

CONTENTS

ITU-T Recommendation J.161

Audio and video codec requirements and usage for the provision of bidirectional audio services over cable television networks using cable modems

1 Scope

This Recommendation 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. This Recommendation also specifies the performance required in client devices to support future IPCablecom codecs. Additionally, this Recommendation describes a suggested methodology for optimal network support for codecs.

1.1 Introduction and overview

The quality of audio and video delivered over the IPCablecom architecture will depend on multiple factors: the end device performance, the network's inherent quality, and the intelligence of the system resource allocation policy. This Recommendation specifies codecs and capabilities supporting audio and video applications, with a particular emphasis on the stringent requirements of IP-based voice communications.

Acceptable voice communications functionality imposes strict delay and packet-loss criteria on IP implementations and will thus stress system resources, particularly if bandwidth becomes congested or saturated. Video applications – while more forgiving to dropped packets and delay – require bandwidth of at least an order of magnitude more than audio applications. The IPCablecom architecture is designed to support both voice and video applications simultaneously.

Speech and video compression are evolving technologies. New algorithms are being enabled as more sophisticated and higher performing processors become available at lower cost. Additionally, the system infrastructure and mechanisms for allocating resources will evolve. Due to this dynamism, the priority in designing the architecture is to define a robust system to accommodate evolving technology without creating a legacy burden.

Therefore, the IPCablecom philosophy is to establish cost-effective envelopes for network and device performance to enable the most appropriate current technology, while allowing upgrades as technology and market needs evolve. To address near-term market needs, this Recommendation also specifies codec and performance mandates to deliver the quality-of-service necessary for launching competitive services.

1.2 Purpose of this Recommendation

The purpose of this Recommendation is to extend the existing IPCablecom v1Codec Recommendation by introducing two new low-bit codecs, [ITU-T T.38] fax relay for reliable fax transmission, [IETF RFC 2833] DTMF Relay for reliable DTMF transmission and metrics to measure voice quality. It is issued to facilitate design and field-testing leading to the manufacturability and interoperability of conforming hardware and software by multiple vendors.

1.3 Phasing of requirements

The codec requirements contained in this Recommendation cover both audio and video multimedia terminals (MTAs) and trunking gateways (media gateway). The term MTA-2 is used to define a terminal supporting video.

In the initial phase of IPCablecom, MTAs are not required to support the MTA-2 requirements as defined in clause 8. MTAs MUST support the requirements for audio terminals as defined in clauses 5, 6, 7, and 9.

Support for video terminals will be REQUIRED in later phases of IPCablecom. All MTA-2s MUST support the requirements defined in clause 8.

2 References

2.1 Normative references

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published. The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

[ITU-T G.107] ITU-T Recommendation G.107 (2005), *The E-model, a computational model for use in transmission planning*.

[ITU-T G.165] ITU-T Recommendation G.165 (1993), *Echo cancellers*.

[ITU-T G.168] ITU-T Recommendation G.168 (2007), *Digital network echo cancellers*.

[ITU-T G.711] ITU-T Recommendation G.711 (1988), *Pulse code modulation (PCM) of voice frequencies*.

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

[ITU-T G.729] ITU-T Recommendation G.729 (2007), *Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-exited linear prediction (CS-ACELP)*.

[ITU-T J.170] ITU-T Recommendation J.170 (2005), *IPCablecom security specification*.

[ITU-T J.171.x] ITU-T Recommendation J.171.x (2005), *IPCablecom trunking gateway control protocol (TGCP)*.

[ITU-T P.56] ITU-T Recommendation P.56 (1993), *Objective measurement of active speech level*.

[ITU-T P.561] ITU-T Recommendation P.561 (2002), *In-service non-intrusive measurement device – Voice service measurements*.

[ITU-T T.38] ITU-T Recommendation T.38 (2007), *Procedures for real-time Group 3 facsimile communication over IP networks*.

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

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

[IETF RFC 3611] IETF RFC 3611 (2003), *RTP Control Protocol Extended Reports (RTCP XR)*.

[IETF RFC 3550] IETF RFC 3550 (2003), *RTP: A Transport Protocol for Real-Time Applications*.

[IETF RFC 3551] IETF RFC 3551 (2003), *RTP Profile for Audio and Video Conferences with Minimal Control*.

[IETF RFC 3951] IETF RFC 3951 (2004), *Internet Low Bit Rate Codec (iLBC)*.

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

[IETF RFC 4298] IETF RFC 4298 (2005), *RTP Payload Format for BroadVoice Speech codecs*.

[ANSI T1.521] ANSI T1.521-1999, *Packet Loss Concealment for Use with ITU-T Recommendation G.711*.

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

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

2.2 Informative references

[ITU-T F.703] ITU-T Recommendation F.703 (2000), *Multimedia conversational services*.

[ITU-T H.243] ITU-T Recommendation H.243 (2005), *Procedures for establishing communication between three or more audiovisual terminals using digital channels up to 1920 kbit/s*.

[ITU-T H.245] ITU-T Recommendation H.245 (2006), *Control protocol for multimedia communication*.

[ITU-T H.261] ITU-T Recommendation H.261 (1993), *Video codec for audiovisual services at p × 64 kbit/s*.

[ITU-T H.263] ITU-T Recommendation H.263 (2005), *Video coding for low bit rate communication*.

[ITU-T H.323] ITU-T Recommendation H.323 (2006), *Packet-based multimedia communications systems*.

[ITU-T H.324] ITU-T Recommendation H.324 (2005), *Terminal for low bit-rate multimedia communication*.

[ITU-T J.160] ITU-T Recommendation J.160 (2005), *Architectural framework for the delivery of time-critical services over cable television networks using cable modems*.

[ITU-T J.163] ITU-T Recommendation J.163 (2005), *Dynamic quality of service for the provision of real-time services over cable television networks using cable modems*.

[ITU-T P.862] 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*.

[ITU-T T.140] ITU-T Recommendation T.140 (1998), *Protocol for multimedia application text conversation*.

[IETF RFC 2327] IETF RFC 2327 (1998), *SDP: Session Description Protocol*.

[IETF RFC 4103] IETF RFC 4103 (2005), *RTP Payload for Text Conversation*.

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

[Telcordia] Telcordia: *LSSGR: Signalling for Analog Interfaces*, Generic Requirements Gr-506-Core Issue 1 June 1996, Revision 1 November 1996, Morristown, New Jersey.

3 Terms and definitions

This Recommendation uses the following terms:

3.1 audio server: An audio server plays informational announcements in IPCablecom network. 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.

3.2 call management server (CMS): Controls the audio connections. Also called a call agent in MGCP/SGCP terminology. This is one example of an application server.

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

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

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

3.6 hybrid fibre/coaxial cable (HFC): An 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.

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

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

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

3.10 media gateway (MG): Provides the bearer circuit interfaces to the PSTN and transcodes the media stream.

3.11 media gateway controller (MGC): The overall controller function of the PSTN gateway. Receives, controls and mediates call-signalling information between the IPCablecom and PSTN.

3.12 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.

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

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

3.15 pulse code modulation (PCM): A 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.

3.16 quality of service (QoS): Guarantees network bandwidth and availability for applications.

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

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

3.19 transit delays: The 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.

3.20 trunk: An analog or digital connection from a circuit switch that carries user media content and may carry voice signalling (M_F, R_2, etc.).

3.21 user datagram protocol (UDP): A 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.

4 Abbreviations, acronyms and conventions

4.1 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

CIF	Common Intermediate Format
CM	DOCSIS Cable Modem
CMS	Call Management Server
CMTS	Cable Modem Termination System
Codec	Coder-DECoder
DOCSIS®	Data-Over-Cable Service Interface Specification
DQoS	Dynamic Quality of Service
DTMF	Dual-tone Multi Frequency (tones)
FEC	Forward Error Correction
HFC	Hybrid Fibre/Coaxial cable
ICMP	Internet Control Message Protocol
IP	Internet Protocol
ISDN	Integrated Services Digital Network
ISP	Internet Service Provider
IVR	Interactive Voice Response System
LSSGR	LATA Switching System Generic Requirements
MG	Media Gateway
MGC	Media Gateway Controller
MTA	Multimedia Terminal Adapter
NCS	Network Call Signalling
NTSC	National Television Standards Committee. (Defines the analog colour television, broadcast standard used today in North America.)

PAL	Phase Alternate Line. (The European colour television format that evolved from the American NTSC standard.)
PCM	Pulse Code Modulation
PSTN	Public Switched Telephone Network
QCIF	Quarter Common Intermediate Format
QoS	Quality of Service
RJ-11	Registered Jack-11
RSVP	Resource Reservation Protocol
RTCP	Real-time Transport Control Protocol
RTP	Real-time Transport Protocol
SDP	Session Description Protocol
TCP	Transmission Control Protocol
UDP	User Datagram Protocol
VAD	Voice Activity Detection
VBR	Variable Bit Rate
VoIP	Voice over IP

4.2 Conventions

Throughout this Recommendation, words that are used to define the significance of particular requirements are capitalized. These words are:

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

5 Background

This clause outlines the IPCablecom 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 Recommendation [ITU-T J.160], 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, delay 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 scores (MOS) 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 delay thresholds have been exceeded, the tradeoff between delay and fidelity becomes an untenable situation.

5.2.2 Delay control

Control of overall delay 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 delay during traversal of an audio signal from capture of the speaker's voice until reception at the receiver's ear. The primary contributors to delay 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 delay in the network are described below.

5.2.2.1 Delay control: Buffering

While network jitter and corresponding buffering increase call delay, 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 delay.

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 transport 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 Delay 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 Delay control: Packet timing optimization

To avoid additional buffering delay, packets MUST 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 mandatory codec is shown in Table 1; further frame sizes for the recommended codecs are listed in Table 2. Default packetization periods are specified in [IETF RFC 3551].

Table 1 – Frame sizes of the mandatory codec

Codec	Frame size (ms)
G.711	0.125

Table 2 – Frame sizes of the recommended codecs

Codec	Frame size (ms)
G.729 Annex E	10
G.728	0.625
iLBC	20
iLBC	30
BV16	5

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 lossless. 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 bit rate codec may be applied, thus reducing the bandwidth required.
- A codec may employ some form of variable bit rate 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 bit rate (VBR) encoding 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 delay 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 upgradability 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 MUST 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 MUST NOT exceed 20 seconds excluding reboot time when downloading new software to the MTA.

6.3 Minimum processing capability

All MTA devices MUST be capable of supporting the equivalent simultaneous execution of codec combinations shown in Table 3. Although this Recommendation does not mandate the support of G.728, G.729 Annex E, iLBC, or BV16, this requirement provides the necessary reserve capacity for additional future codecs to be provisioned (configured and downloaded) on the MTA. The MTA MUST support T.38 fax relay on all ports simultaneously. Media gateway MUST 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 MUST be supported on all connections simultaneously by both MTA and media gateway.

Table 3 – 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				

Table 3 – MTA processing capability

Maximum ports supported by MTA	G.711 ports	iLBC ports	BV16 ports	G.728 ports	G.729E ports
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 MUST 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 BV16). Although this Recommendation does not mandate G.728, G.729 Annex E, iLBC, or BV16, 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, 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 this Recommendation, 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 autodialling 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) MUST successfully pass DTMF tone transmissions in band via RFC 2833 telephone events (clause 7.1.9) subject to a successful negotiation. When negotiation is unsuccessful, e.g., due to interworking with older non-RFC 2833-capable endpoints, DTMF tone transmissions MUST be passed in the regular audio stream using the voice codec by MTAs and MGs.

The capability described above MUST 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 small office (or home office) 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 MUST 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 2100 Hz tone with periodic phase reversals per [ITU-T G.165] and [ITU-T G.168]. After the device session has completed, echo compensation MUST 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 [ITU-T T.38]. MTAs and media gateways MUST support T.38 fax relay as defined in clause 7.1.8.

MTAs and MGs MUST detect the T.30 fax preamble (V.21 flags) and CNG (calling fax tone). The detection of CNG MUST 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 kbit/s max) in T.38 implementations not capable of supporting Version 3 (ITU-T Rec. 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 kbit/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 MUST apply appropriate signal discrimination to minimize the chance that a voice call could inadvertently be switched to T.38 fax relay.

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 MUST be provided in IPCablecom MTA and gateway devices to mitigate the effects of line echo. This echo canceller MUST 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 echo canceller MUST comply with either [ITU-T G.165] or [ITU-T G.168].

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 ms or less). For PSTN gateway applications, the echo canceller length is typically much longer (32 ms 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 hearing-impaired people and people 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 November 2000, adopted [ITU-T V.18] to begin alleviating this problem. [ITU-T 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 voiceband modems, any system designed to support them would need to be able to pass DTMF and voiceband 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 MUST support detection of ITU-T V.18 hearing-impaired tones, including V.18 Annex A. Upon detection of a V.18 signal, the MTA MUST 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 MUST notify the CMS of the modem tone event, if this event is in the Requested Events list. The MTA MUST disable echo cancellation for the remainder of the session when phase reversals are present in the answer tone, in accordance with [ITU-T G.168].

Upon detection of a V.18 signal, the codec at each end MUST be switched to a codec that supports transmission of V.18 tones for the remainder of the session, for example the G.711 codec. The endpoints MUST 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 MUST 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 MUST be supported.

7.1.7 Packet loss concealment

All media gateways and media terminal adapters MUST 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 [IETF RFC 3952] SHOULD be used for iLBC and the packet loss concealment method for BV16, as defined and included in [IETF RFC 4298] SHOULD be used for BV16). For G.711, the method defined in [ANSI T1.521] SHOULD be used.

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.

[ITU-T T.38] is a widely recognized standard for fax relay. The first version for [ITU-T T.38] is version 1 and the majority of implementations are compatible with this version, while later implementations are also required to interoperate with version 1. MTAs and media gateways MUST support version 1 of [ITU-T T.38] in order to ensure interoperability with existing T.38 implementations. In addition, a MTA or a media gateway MAY support any other version of [ITU-T T.38]. All client devices (MTAs and media gateways) MUST support the V.27 *ter*, V.29, V.17 modem protocols for page transmission within the T.38 implementation to allow transfer rates up to 14'400 bit/s. 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 the implementation supports T.38 version 4, then it MAY use T.38 version 4 mode to transport super G3 fax transmissions.

7.1.8.1 T.38 over UDPTL

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

[ITU-T 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 ITU-T Rec. J.162.

To control the T.38 UDPTL session, the FXR package will be used and all endpoints MUST support this package as described in ITU-T Rec. J.162.

The MTA MUST be prepared to receive a T.38 UDPTL fax packet of at least 160 bytes in the downstream. This is based on a 40-ms packetization period and a 14'400 bit/s data rate. It includes the UDPTL datagram without the IP and UDP headers.

For DQoS considerations, T.38 fax packets SHOULD use the same port used by the voice packets for the connection. In addition, the MTA MUST 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 (10/20/30 ms). Similarly, the MG MUST send T.38 fax packets at a default 20 ms packetization period in the upstream unless directed by the MGC via the packetization period to use a different packet rate (10/20/30 ms).

Table 5 shows the DQoS flowspec parameters for 10/20/30 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 MUST 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 [ITU-T T.38] is currently out of scope.

7.1.9 DTMF relay

[IETF RFC 2833] 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, [IETF RFC 2833] 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 [IETF RFC 2833] 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 [IETF RFC 2833] 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 MUST support transmission and reception of RFC 2833 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 MUST 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 MUST be secured through the IPCablecom bearer encryption and authentication mechanisms defined in [ITU-T J.170], if these are active on a call. MTAs and MGs MUST support the mandatory security options listed in [ITU-T J.170] for DTMF relay and additionally, if the optional encryption algorithms are supported for audio codecs, then these MUST also be supported for DTMF relay.

[IETF RFC 2833] references ITU-T Rec. Q.24 in defining the minimum DTMF tone duration of 40 ms. Additionally, ITU-T Rec. Q.24 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 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 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.

[IETF RFC 2833] 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 MUST 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 MUST NOT report DTMF digits when their duration is less than 23 ms.

An IPCablecom endpoint MUST 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 MUST be equal to the same packetization time as the selected audio codec. Therefore the repetition rate of RFC 2833 packets has the same range as packetization intervals, i.e., 10, 20 and 30 ms.

In accordance with [IETF RFC 2833], unless a mutually exclusive event (detection of new DTMF digit) occurs, the final packet of each event MUST 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 MUST be aborted and instead the new DTMF telephone event reported using the regular packetization interval.

Upon receipt of any telephone-event packet, IPCablecom endpoints MUST 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 on the TDM interface, this does not constitute a signalling event and the call agent or media gateway controller MUST NOT be informed of this.

[IETF RFC 2833] 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 MUST use the continuous method of play out.

[IETF RFC 2833] 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 MUST stop sending audio and replace audio packets with RFC 2833 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 MUST 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 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 MUST 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 packets MUST continue to be suppressed during the redundant transmission of the end-of-event packets.

7.2 Mandatory codec

G.711 (both μ-law and A-law versions) MUST 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 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, MTAs and MGs SHOULD also support at least one of the following codecs.

7.3.1 G.728

G.728 SHOULD be supported in all MTAs. IPCablecom has as a mandate to provide toll or superior voice quality. G.728 is a mid-bit rate (16 kbit/s), high-quality solution. Recommending a codec in this range 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.

7.3.2 Annex E of G.729

Annex E of G.729 SHOULD be supported in all MTAs. IPCablecom has as a mandate to provide toll or superior voice quality. G.729E is a mid-bit rate (11.8 kbit/s), high-quality solution. Recommending a codec in this range 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.

7.3.3 iLBC

iLBC SHOULD be supported in all MTAs and MGs. IPCablecom has as a mandate to provide toll or superior voice quality. iLBC is a mid-bit rate (13.3 kbit/s and 15.2 kbit/s), high-quality solution. When iLBC is supported, both the 20 ms and 30 ms frame size modes MUST be supported. Recommending a codec in this range 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. In addition, it provides DTMF pass through. Note that [IETF RFC 3951] and [IETF RFC 3952] are classified as experimental.

7.3.4 BV16

BroadVoice16 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-bit rate, high-quality solution. Recommending a codec in this range 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-based telecommunication networks.

7.4 Optional features

7.4.1 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 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.

7.4.2 Optional codecs

The G.726 and G.729 Annex A codecs are optional. A vendor MAY supply any codecs not described herein.

7.4.3 Voice activity detection (VAD)

A vendor MAY employ VAD to reduce bandwidth consumption. If employed, this capability MUST be optional, allowing disabling. Some codecs have associated VAD implementations (e.g., G.729B), while many others do not (e.g., G.711 and G.728). In the latter cases, the VAD implementation MUST adhere to the [IMTC] Voice-Over-IP Forum Service Interoperability Implementation Agreement 1.0.

7.4.4 Real time text feature support

Text used as a real time medium is an important part of rich media communication. The text is intended to be entered by human users from a keyboard, handwriting recognition, voice recognition or any other input method. The rate of character entry is usually at a level of a few characters per second or less. In general, only one or a few new characters are expected to be transmitted with each packet.

Real time text gives an opportunity to create and maintain the feeling of contact in a conversation. Commonly text will be combined with an audio call, so that depending on the situation, the real time text medium can be used to carry small bits of the conversation, or, the whole conversation in text.

When text is combined with audio, the real-time communication may be established as described in [ITU-T F.703], [ITU-T T.140] and [IETF RFC 4103].

Note that the establishment of real time text communications combined with audio may have implications beyond this codec Recommendation, for example QoS and signalling.

Additional information concerning accessibility may be found at
http://www.itu.int/ITU-T/studygroups/com16/accessibility/docs/checklist.

7.5 Session description of codecs

Session description 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) (ITU-T Rec. J.162). 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, [IETF RFC 3551].

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 [IETF RFC 3551] are dynamically bound by using a dynamic payload type from the range 96-127, as defined in [IETF RFC 2327], 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 8000 samples/s.

Codecs defined in this Recommendation MUST be encoded with the following string names in the rtpmap parameter:

Table 4 – 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 kbit/s	G726-16	G726-16/8000
G.726 at 24 kbit/s	G726-24	G726-24/8000
G.726 at 32 kbit/s	G726-32	G726-32/8000
G.726 at 40 kbit/s	G726-40	G726-40/8000
G.728	G728	G728/8000
G.729A	G729	G729/8000
G.729E	G729E	G729E/8000
RFC 2833 DTMF	telephone-event	telephone-event/8000
NOTE: Mandatory codec – G.711 (μ-law and A-law) and RFC 2833 DTMF. Recommended codecs – G.728, G.729 Annex E, iLBC and BV16. Optional codecs (for informational purposes only) – G.726 and G.729A.		

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), Table 5 describes the mapping that MUST 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 5 do not include any bandwidth that may be required for media security (authentication, 2- or 4-byte value as outlined in the security Recommendation), 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 [ITU-T J.163]) MUST be according to Table 5.

Table 5 – Mapping of session description parameters to RSVP flowspec

Parameters from session description			Flowspec parameters		Comments
RTP/AVP code	**Rtpmap**	**Ptime (ms)**	**Values b, m, M[1]**	**Values r, p[2]**	
0	<none>	10	120 bytes	12'000 bytes/s	G.711 µ-law using the payload type defined by IETF
0	<none>	20	200 bytes	10'000 bytes/s	
0	<none>	30	280 bytes	9334 bytes/s	
96-127	PCMU/8000	10	120 bytes	12'000 bytes/s	G.711 µ-law PCM, 64 kbit/s, default codec
96-127	PCMU/8000	20	200 bytes	10'000 bytes/s	
96-127	PCMU/8000	30	280 bytes	9334 bytes/s	
8	<none>	10	120 bytes	12'000 bytes/s	G.711 A-law using the payload type defined by IETF
8	<none>	20	200 bytes	10'000 bytes/s	
8	<none>	30	280 bytes	9334 bytes/s	
96-127	PCMA/8000	10	120 bytes	12'000 bytes/s	G.711 A-law PCM, 64 kbit/s, default codec
96-127	PCMA/8000	20	200 bytes	10'000 bytes/s	
96-127	PCMA/8000	30	280 bytes	9334 bytes/s	
96-127	iLBC/8000	20	78 bytes	3900 bytes/s	iLBC, FB-LPC, 15.2 kbit/s, 20 ms frame size with 5 ms lookahead; 13.3 kbit/s, 30 ms frame with 10 ms lookahead
96-127	iLBC/8000	30	90 bytes	3000 bytes/s	
96-127	BV16/8000	10	60 bytes	6000 bytes/s	BV16 (narrow-band), 16 kbit/s
96-127	BV16/8000	20	80 bytes	4000 bytes/s	
96-127	BV16/8000	30	100 bytes	3334 bytes/s	
96-127	G726-16/8000	10	60 bytes	6000 bytes/s	
96-127	G726-16/8000	20	80 bytes	4000 bytes/s	
96-127	G726-16/8000	30	100 bytes	3334 bytes/s	
96-127	G726-24/8000	10	70 bytes	7000 bytes/s	
96-127	G726-24/8000	20	100 bytes	5000 bytes/s	
96-127	G726-24/8000	30	130 bytes	4334 bytes/s	

Table 5 – Mapping of session description parameters to RSVP flowspec

Parameters from session description			Flowspec parameters		Comments
RTP/AVP code	**Rtpmap**	**Ptime (ms)**	**Values b, m, M**[1]	**Values r, p**[2]	
2	<none>	10	80 bytes	8000 bytes/s	G.726-32, identical to G.721, which is assigned payload type 2 by IETF
2	<none>	20	120 bytes	6000 bytes/s	
2	<none>	30	160 bytes	5334 bytes/s	
96-127	G726-32/8000	10	80 bytes	8000 bytes/s	
96-127	G726-32/8000	20	120 bytes	6000 bytes/s	
96-127	G726-32/8000	30	160 bytes	5334 bytes/s	
96-127	G726-40/8000	10	90 bytes	9000 bytes/s	
96-127	G726-40/8000	20	140 bytes	7000 bytes/s	
96-127	G726-40/8000	30	190 bytes	6334 bytes/s	
15	<none>	10	60 bytes	6000 bytes/s	G.728, assigned payload type 15 by IETF
15	<none>	20	80 bytes	4000 bytes/s	
15	<none>	30	100 bytes	3334 bytes/s	
96-127	G728/8000	10	60 bytes	6000 bytes/s	G.728, LD-CELP, 16 kbit/s
96-127	G728/8000	20	80 bytes	4000 bytes/s	
96-127	G728/8000	30	100 bytes	3334 bytes/s	
18	<none>	10	50 bytes	5000 bytes/s	G.729A, identical to G.729, assigned payload type 18 by IETF
18	<none>	20	60 bytes	3000 bytes/s	
18	<none>	30	70 bytes	2334 bytes/s	
96-127	G729/8000	10	50 bytes	5000 bytes/s	G.729A, CS-ACELP, 8 kbit/s, 10 ms frame size with 5 ms lookahead
96-127	G729/8000	20	60 bytes	3000 bytes/s	
96-127	G729/8000	30	70 bytes	2334 bytes/s	
96-127	G729E/8000	10	55 bytes	5500 bytes/s	G.729E, CS-ACELP, 11.8 kbit/s, 10 ms frame size with 5 ms lookahead
96-127	G729E/8000	20	70 bytes	3500 bytes/s	
96-127	G729E/8000	30	85 bytes	2834 bytes/s	
N/A	N/A	10	80 bytes	800 bytes/s	T.38 fax relay packets (with T.4 redundancy level 1, T30 redundancy level 4)
N/A	N/A	20	116 bytes	500 bytes/s	
N/A	N/A	30	152 bytes	567 bytes/s	

NOTE:

Mandatory codec – G.711 (μ-law and A-law).

Recommended codecs – G.728, G.729 Annex E, iLBC and BV16.

Optional codecs (for informational purposes only) – G.726 and G.729A.

[1] *b* is bucket depth (bytes). *m* is minimum policed unit (bytes). *M* is maximum datagram size (bytes).

[2] *r* is bucket rate (bytes/s). *p* is peak rate (bytes/s).

7.5.1 iLBC session description

Parameters are mapped to SDP in a standard way. When conveying information by SDP, the encoding name MUST be "iLBC" (the same as the MIME subtype in [IETF RFC 3951]).

If 20 ms frame size mode is used, local iLBC encoder MUST 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 MUST be applied.

7.5.2 BV16 session description

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

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
```

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. [ITU-T H.261] and [ITU-T H.263] (as well as [ITU-T H.245], or a functionally equivalent specification) are the basis and reference for this Recommendation; 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 bit rate may be variable. In [ITU-T H.263], no constraints on the video bit rate are given; the terminal or the network, as determined by the CMS or gatekeeper, provides constraints.

For reasons of interoperability, all IPCablecom MTA-2 terminals providing video communications MUST be capable of encoding and decoding video according to [ITU-T H.261]. This will permit video communication without the transcoding of video with terminals across the other networks, such as H.320 terminals across an ISDN network or an H.324 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 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 MUST also be capable of encoding and decoding video according to [ITU-T H.263]. The most important improvement in [ITU-T H.263] is the advancement in motion estimation accuracy to a half-pixel, yielding a lower bit-per-picture requirement at a given bit rate. This, as well as several other advancements in the H.263 baseline codec and its annexes, result in a higher frame rate and/or resolution at a given bit rate versus [ITU-T H.261].

8.4 Video format requirements

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

> "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 [ITU-T H.261]; SQCIF, 4CIF and 16CIF are defined in [ITU-T H.263].

Table 6 – 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	1408	1152	704	576

An MTA-2 MUST 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 bit rates ranging from 128 kbit/s to 768 kbit/s. 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 multipoint 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 bit rate, picture format, and algorithm options, which can be accepted by the decoder, MUST 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 MUST be capable of accepting mandatory video bitstreams that do not make use of that option.

MTA-2 terminals MUST 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 [ITU-T H.263], when each video logical channel is opened, the maximum operating mode to be used on that channel MUST be signalled to the receiver. The maximum mode signalled includes maximum picture format, algorithm options, maximum codec bit rate, etc., as defined in [ITU-T 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 MUST 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 [ITU-T H.263]) of each of the annexes follow. In order to guide vendor development and to encourage the highest common denominator of video quality possible employing [ITU-T H.263], 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 bit rates) 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 Specification

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 bit rate, 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 Table 7. Also listed (for purposes of comparison only) are the three levels of preferred mode support described in Appendix II of [ITU-T H.263].

Table 7 – 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 [ITU-T 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 Recommendation, the precise signalling protocol for all client devices has not been specified, but the following discussion demonstrates the necessary signals, whatever the protocol.

[ITU-T H.245] 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 in Table 8 for receiving and transmitting MTA-2s. The mandatory commands (or their functional equivalents) MUST be implemented in the IPCablecom signalling system.

Table 8 – 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
Conference Commands		
Terminal List Request	O	O
Broadcast Me	O	O
Cancel Broadcast Me	O	O

Table 8 – H.245 Commands that are applicable to IPCablecom

Message	Receiving MTA status	Transmitting MTA status
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
M Mandatory O Optional		

9 RTP and RTCP usage

9.1 RTP requirements

The voice and fax/modem pass-through media flows MUST be transported using IETF real-time transport protocol (RTP) and real-time transport control protocol (RTCP) as defined in [IETF RFC 3550] and [IETF RFC 3551]. All IPCablecom devices supporting RTP (e.g., MTAs, trunking gateways, audio servers) MUST support RTCP as defined in [IETF RFC 3550] and [IETF RFC 3551] 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 [IETF RFC 3550] and [IETF RFC 3551], 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 MUST 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 MUST send RTCP messages, as described in [IETF RFC 3550] and [IETF RFC 3551] 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 MUST start transmitting RTCP messages if it receives an RTCP message. Once started, the endpoint MUST 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 MUST receive RTCP messages, if sent by the remote communication peers. IPCablecom endpoints MUST NOT require them. That is, call state in general and RTP flows in particular MUST 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 MUST 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 MUST 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 MUST 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 MUST NOT contain identity information (see definition below); CNAME field MUST be a cryptographically-random value generated by the endpoint in such a manner that endpoint identity is not compromised and MUST change on a per-session basis; NAME, EMAIL, PHONE, LOC objects SHOULD NOT be sent and, if sent, MUST NOT contain identity information. This requirement is intended to satisfy the requirements of [IETF RFC 3550] 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 Recommendation [ITU-T J.170].

SR (sender report): MUST be sent by IPCablecom endpoints transmitting RTP packets (as described in [IETF RFC 3551]) 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): MUST be sent with report blocks if receiving but not sending RTP packets (as described in [IETF RFC 3550]) and MUST 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 MUST NOT contain identity info. Endpoints MUST ignore and silently discard APP messages with unrecognized contents.

BYE (goodbye): MUST be sent upon RTP connection deletion or when renegotiating SSRC upon collision detection and resolution (see below). Endpoints MUST 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 MUST NOT use BYE messages to indicate or detect any call progress condition. For example, endpoints MUST NOT tear down RTP flows based on BYE, but MUST update RTCP/RTP state as per [IETF RFC 3550]. 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 MUST NOT conform to the security requirements described in the RTP/RTCP RFC and drafts. Instead, IPCablecom endpoints MUST implement RTP and RTCP security as specified in the IPCablecom security Recommendation [ITU-T J.170].

9.2.3 Extended RTCP reports

The RTCP XR VoIP metrics report block as defined in [IETF RFC 3611] MUST be sent by endpoints if negotiated on a given connection as defined in IPCablecom network-based call signalling protocol Recommendation (ITU-T Rec. J.162 and trunking gateway control protocol Recommendation [ITU-T J.171.x]). IPCablecom endpoints MAY send other RTCP XR payload types. IPCablecom endpoints that are capable of sending RTCP XR reports MUST 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 [IETF RFC 3611] 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 analysers, 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 MUST exchange RTCP XR VoIP metrics reports during active RTP sessions if negotiated and MUST concatenate RTCP XR payloads with RTCP SR and RR payloads, following rules for transmission intervals per [IETF RFC 3550].

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

9.2.3.1.2 Definition of metrics related to packet loss and discard

The VoIP metrics [IETF RFC 3611] 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 MUST be set to 16 within IPCablecom systems.

An IPCablecom endpoint when using RTCP XR MUST provide these parameters as defined in Table 9.

Table 9 – 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

9.2.3.1.3 Definition of metrics related to delay

The VoIP metrics payload includes two delay metrics [IETF RFC 3611]. The round trip delay is the delay between RTP interfaces, as typically measured using RTCP sender report (SR) or receiver report (RR) [IETF RFC 3550]. The end system delay incorporates the vocoder encoding and decoding delay, the packetization delay, and the current nominal delay due to the jitter buffer.

An IPCablecom endpoint when using RTCP XR MUST provide the parameters as defined in Table 10. (Note that this requires an SR or RR exchange prior to the inclusion of an XR payload into an RTCP message.)

Table 10 – 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

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.

An IPCablecom endpoint when using RTCP XR MUST provide Signal Level and Noise level as defined in Table 11.

An IPCablecom endpoint equipped with an echo canceller and when using RTCP XR MUST provide the Residual Echo Return Loss metric as defined in Table 11.

Table 11 – Metrics due to signal

Metric	Description	Range
Signal Level	RMS Signal level during active speech periods (dBm0). As defined in P.56 and P.561.	–30 to +3
Noise Level	RMS Noise level during silence periods (dBm0). As defined in P.56 and P.561.	–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 G.168.	0 to 80

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 (3611). 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 of [ITU-T G.107]) or in terms of mean opinion score (MOS).

The maximum range of an R factor is 0-100 for narrow-band voice transmission. Note, however, for wideband transmission the upper range can be greater than 100. The R factor defined in the ITU-T 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:

$$R = Ro - Is - Id - Ie\text{-}eff + A$$

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 G.107] defines an equation for converting an R factor into a MOS score; note however that this produces MOS scores slightly higher than those typically reported from subjective tests.

Table 12 – 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 (×10)	10 to 50
MOS-CQ	Estimated conversational quality MOS (×10)	10 to 50

An IPCablecom endpoint when using RTCP XR MUST provide the R factor, MOS-LQ and MOS-CQ metrics and MAY provide an external R factor.

An IPCablecom endpoint when using RTCP XR MUST calculate R factors using [ITU-T G.107] at a minimum.

An IPCablecom endpoint when using RTCP XR MUST 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 MUST 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 the E-model default value for SLR MUST be assumed. For more information, refer to [ITU-T G.107].

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 the E-model default value for TELR MUST be assumed. For more information, refer to [ITU-T G.107].

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 [ITU-T G.107].*

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 [ITU-T G.107].

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 [ITU-T G.107].

An IPCablecom endpoint when using RTCP XR MUST calculate the Ie-eff parameter using the function defined in [ITU-T G.107]. However, the IPCablecom endpoint MUST use the Ie and Bpl parameters defined in Table 13 for the vocoder and PLC combinations listed.

Table 13 – Ie and Bpl parameters for IPCablecom vocoders

Vocoder	Bit rate	PLC	Ideal R	Ideal MOS	Ie	Bpl
G.711 A/U	64 kbit/s	Appendix 1 ANSI T1.521	93	4.4	0	34
G.728 10 ms	16 kbit/s	Per G.728 Annex I	89	4.1	7	17
G.728 20 ms	16 kbit/s	Per G.728 Annex I	89	4.1	7	15
G.729 Annex E 10 ms	11.8 kbit/s	Per G.729	88	4.1	4	20
G.729 Annex E 20 ms	11.8 kbit/s	Per G.729	88	4.1	4	19
iLBC 20 ms	15.2 kbit/s	Per iLBC code in RFC 3951	80	3.9	10	34
iLBC 30 ms	13.3 kbit/s	Per iLBC code in RFC 3951	78	3.8	12	27
BV16 10 ms	16 kbit/s	Per {BV16}	88	4.2	5	25
BV16 20 ms	16 kbit/s	Per {BV16}	88	4.2	5	23

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

An IPCablecom endpoint when using RTCP XR MUST calculate MOS-CQ using the R to MOS mapping function defined in [ITU-T G.107] 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 P.862] 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 [ITU-T 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 score with other codecs with known Ie factor.

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

Estimated PESQ score = 1 + 0.033.R_{adj} + R_{adj}(100-R_{adj})(R_{adj}–60)*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 score, gives the closest match to the measured P.862 score. Alternatively, the Bpl value may be obtained by comparing the P.862 score curve with other codecs with known Bpl factor.

d) It is generally advisable to compare the curve of estimated MOS score (derived per [ITU-T G.107]) 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 14 describe some key configuration parameters of the IPCablecom endpoint, that are useful in monitoring service quality and identifying some types of configuration related problems.

An IPCablecom endpoint when using RTCP XR MUST provide values to all parameters as defined in Table 14.

Table 14 – Parameters related to endpoint configuration

Metric	Description	Range
PLC Type	Type of packet loss concealment algorithm	UnspecifiedDisabled EnhancedStandard
Jitter Buffer Type	Type of jitter buffer (fixed or adaptive)	UnknownReserved Non-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

Appendix I

Codec comparison tables

(This appendix does not form an integral part of this Recommendation)

Tables I.1 to I.3 summarize standard speech coder characteristics. Some of the data in the three tables are obtained from "Current Methods of Speech Coding".

Table I.1 – ITU-T IETF and SCTE Speech Coders

Standards body	ITU-T	ITU-T	ITU-T	ITU-T	ITU-T	ITU-T	ITU-T	ITU-T	IETF[1]	SCTE
Recommendation	G.711	G.726	G.728	G.729	G.729A	G.729D	G.729E	G.723.1	iLBC	BV16
Coder Type	Companded PCM	ADPCM	LD-CELP	CS-ACELP	CS-ACELP	CS-ACELP	CS-ACELP	MPC-MLQ & ACELP	FB-LPC	TSNFC
Dates	1972	1990	1992/9	1996	1996	1998	1998	1996	2002	2003
Bit rate	64 kbit/s	16-40 kbit/s	16 kbit/s	8 kbit/s	8 kbit/s	6.4 kbit/s	11.8 kbit/s	6.3 kbit/s & 5.3 kbit/s	15.2 kbit/s & 13.3 kbit/s	16 kbit/s
Peak Quality[2]	Toll	≤ Toll	Toll	Toll	Toll	< Toll	Toll	≤ Toll	Toll	Toll
Background Noise[3]	Toll	≤ Toll	Toll	≤ Toll	≤ Toll	< Toll	Toll	≤ Toll	Toll	Toll
Tandem[4]	Toll	Toll	Toll	< Toll	< Toll	< Toll	Toll	< Toll	< Toll	Toll
Frame Erasures[5]	No mechanism	No mechanism	3%	3%	3%	3%	3%	3%	7% & 5%	5%
Complexity (MIPS)[6]	~0.35	~12	~36	~22	~13	~20	~27	~19	~15 & ~18	~12
RAM (kword)[7]	~0.01	~0.15	~2.20	~2.6	~2.6	~2.6	~2.6	~2.1	~4	~2
Frame Size	0.125 ms	0.125 ms	0.625 ms	10 ms	10 ms	10 ms	10 ms	30 ms	20 ms & 30 ms	5 ms
Look Ahead	0	0	0	5 ms	5 ms	5 ms	5 ms	7.5 ms	5 ms & 10 ms	0
Codec Delay[8]	0.25 ms	0.25 ms	1.25 ms	25 ms	25 ms	25 ms	25 ms	67.5 ms	45 ms & 70 ms	10 ms

1 The codec description is in RFC 3951.

2 Peak quality means clean input speech and clear channel for single encoding.

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

4 Tandems refer to the performance of the coder for multiple asynchronous encodings. Toll quality is defined as the performance of 32 kbit/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.

5 Frame erasures refer to the rate at which the MOS score is approximately 0.5 MOS worse than the peak quality for that coder.

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.

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.

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 must 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 I.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
Bit rate	7.95 kbit/s	7.4 kbit/s	0.8-8.0 kbit/s	0.8-8.55 kbit/s	0.8-13.2 kbit/s	13 kbit/s	5.6 kbit/s	12.2 kbit/s
Peak Quality[1]	= GSM-(FR)	Toll	= GSM-(FR)	Toll	Toll	< Toll	= GSM-(FR)	Toll
Background Noise[2]	<< Toll	< Toll	<< Toll	< Toll	Toll	< Toll	< GSM-(FR)	Toll
Tandem[3]	<< Toll	< Toll	<< Toll	< Toll	Toll	<< Toll	< GSM-(FR)	Toll
Frame Erasures[4]	3%	3%	3%	3%	3%	3%	3%	3%
Complexity (MIPS)[5]	~12	~15	~18	~25	~22	~5	~24	~18
RAM (kword)[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[7]	45 ms	45 ms	45 ms	45 ms	45 ms	40 ms	44.4 ms	40 ms

[1] Peak quality means clean input speech and clear channel for single encoding.

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

[3] Tandems refer to the performance of the coder for multiple asynchronous encodings. Toll quality is defined as the performance of 32 kbit/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.

[4] Frame erasures refer to the rate at which the MOS score is approximately 0.5 MOS worse than the peak quality for that coder.

[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.

[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.

[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 must 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.

[ITU-T G.729] was finalized in 1996 originally by ITU-T to be a toll quality 8 kbit/s standard. In that year, ITU-T 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 kbit/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 30th-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 I.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 that G.729E is shown at a byte-boundary 12 kbit/s, which includes the 2 bits/frame not currently defined. Variable bit rate VAD implementations for each codec are not listed.

Table I.3 – Bandwidth attributes of codecs

Codec	Bit rate (kbit/s)	Byte/10 ms	Frm/Pkt	Byte/Pkt	Pkt/s	Byte/s	kbit/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	9333	75
G.726.16-10 ms	16	20	1	60	100	6000	48
G.726.16-20 ms	16	20	2	80	50	4000	32
G.726.16-30 ms	16	20	3	100	33.3	3333	27
G.726.24-10 ms	24	30	1	70	100	7000	56
G.726.24-20 ms	24	30	2	100	50	5000	40
G.726.24-30 ms	24	30	3	130	33.3	4333	35
G.726.32-10 ms	32	40	1	80	100	8000	64
G.726.32-20 ms	32	40	2	120	50	6000	48
G.726.32-30 ms	32	40	3	160	33.3	5333	43
G.726.40-10 ms	40	50	1	90	100	9000	72
G.726.40-20 ms	40	50	2	140	50	7000	56
G.726.40-30 ms	40	50	3	190	33.3	6333	51
G.728-10 ms	16	20	1	60	100	6000	48
G.728-20 ms	16	20	2	80	50	4000	32
G.728-30 ms	16	20	3	100	33.3	3333	27
G.729A-10 ms	8	10	1	50	100	5000	40
G.729A-20 ms	8	10	2	60	50	3000	24
G.729A-30 ms	8	10	3	70	33.3	2333	19
G.729E-10 ms	12	15	1	55	100	5500	44
G.729E-20 ms	12	15	2	70	50	3500	28
G.729E-30 ms	12	15	3	85	33.3	2833	23
iLBC-20 ms	15.2	19	1	78	50	3900	31
iLBC-30 ms	13.3	16.67	1	90	33.3	3000	24
BV16-10 ms	16	20	2	60	100	6000	48
BV16-20 ms	16	20	4	80	50	4000	32
BV16-30 ms	16	20	6	100	33.33	3333	26.7

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
Series D	General tariff principles
Series E	Overall network operation, telephone service, service operation and human factors
Series F	Non-telephone telecommunication services
Series G	Transmission systems and media, digital systems and networks
Series H	Audiovisual and multimedia systems
Series I	Integrated services digital network
Series J	**Cable networks and transmission of television, sound programme and other multimedia signals**
Series K	Protection against interference
Series L	Construction, installation and protection of cables and other elements of outside plant
Series M	Telecommunication management, including TMN and network maintenance
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	Telephone transmission quality, telephone installations, local line networks
Series Q	Switching and signalling
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminals for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	Data networks, open system communications and security
Series Y	Global information infrastructure, Internet protocol aspects and next-generation networks
Series Z	Languages and general software aspects for telecommunication systems

Printed in Switzerland
Geneva, 2008