International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

J.361
(11/2006)

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

IPCablecom

IPCablecom2 codec and media

ITU-T Recommendation J.361

ITU-T

ITU-T Recommendation J.361

IPCablecom2 codec and media

Summary

This new IPCablecom2 Recommendation addresses interfaces for audio, video and text communications between user equipment (UE), media gateways (MG) and other network elements, such as media servers, that process media. Specifically, it identifies the audio and video codecs and other features necessary to provide the highest quality and the most resource-efficient media streams to the customer.

Source

ITU-T Recommendation J.361 was approved on 29 November 2006 by ITU-T Study Group 9 (2005-2008) under the ITU-T Recommendation A.8 procedure.

It includes the modifications introduced by Amendment 1 approved on 29 July 2007.

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 not 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.361

IPCablecom2 codec and media

1 Scope and introduction

1.1 Scope

This Recommendation addresses interfaces for audio, video and text communications between user equipment (UE), media gateways (MG) and other network elements, such as media servers that process media. Specifically, it identifies the audio and video codecs and other features necessary to provide the highest quality and the most resource-efficient media streams to the customer.

1.2 Introduction

1.2.1 Background

IPCablecom2 defines a modular architecture and a set of interoperable interfaces that leverage emerging communications technologies, such as SIP, to support the rapid introduction of new IP-based communications and streaming services onto the cable network. A modular approach allows operators to deploy network capabilities as required by their specific service offerings, while maintaining interoperability across a variety of devices from multiple suppliers. Examples of the service capabilities include:

- Enhanced residential VoIP and IP video communications – Capabilities such as wideband audio and video telephony plus click-to-dial type call processing based on presence, device capability, and identity.
- Cross platform feature integration – Capabilities such as caller-id and video telephony display on the television, and call treatment from the television.
- Mobility services and integration with cellular and wireless networks – Capabilities such as call hand-off and roaming between IPCablecom VoIP over wireless LAN and cellular networks.
- Multimedia applications – Capabilities such as QoS-enabled audio and video streaming, and real-time conversational text.

The network capabilities required to support this variety of services encompass protocol support, media support, architecture and network element types, security, bandwidth management, and network management. As with previous IPCablecom Recommendations, IPCablecom2 leverages existing open standards wherever possible.

IPCablecom2 is based on the IP multimedia subsystem (IMS) as defined by the 3rd generation partnership Project (3GPP).

Figure 1-1 below illustrates the functional components that are included in the IPCablecom2 architecture. Refer to the IPCablecom2 architecture framework Recommendation for more detail [b-ITU-T J.360].

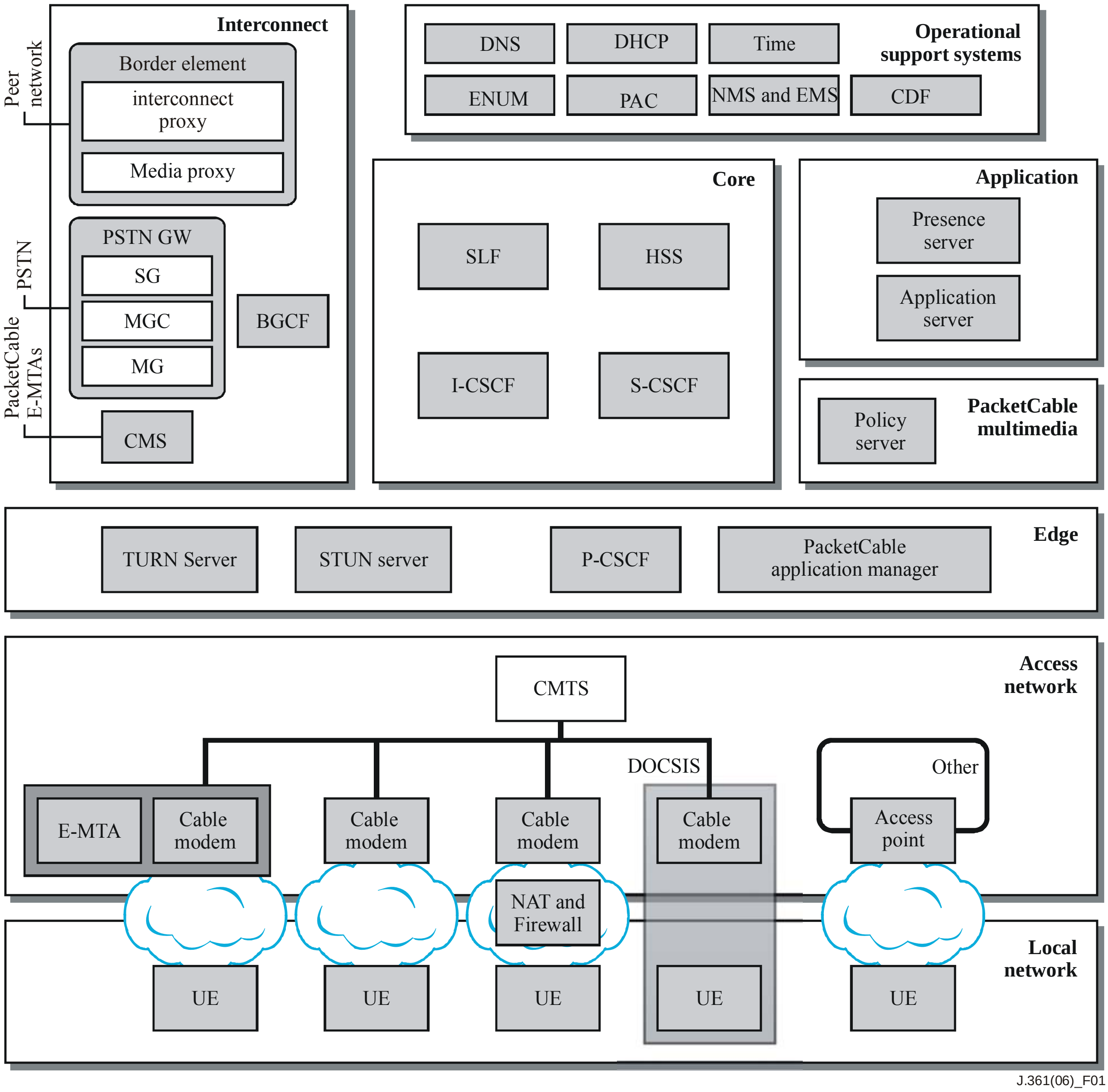

Figure 1-1 – IPCablecom reference architecture

Concerning media support, the subject of this Recommendation, the quality of audio, video and text delivered over the IPCablecom2 architecture depends on multiple factors starting with the inherent capabilities and performance of the end devices, the network's performance and quality, and the intelligence of the network resource allocation. To assure interoperability for both "on-net", including different VoIP network types such as cable-Wi-Fi networks, and "off-net" connections, this Recommendation defines codecs and capabilities for supporting narrow-band and wideband audio applications, video and real-time conversational text applications, with emphasis on the stringent requirements of IP-based voice, video and text communications.

Acceptable two-way voice and video communications and real-time text conversation imposes strict latency and packet-loss criteria on IP implementations and will thus stress system resources, particularly if bandwidth becomes congested or saturated. Entertainment-quality audio and video streaming applications, while more tolerant of latency, still impose strict packet-loss requirements and generally require more bandwidth than two-way communications applications. The IPCablecom architecture is designed to support both types of applications simultaneously.

Audio compression 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 IPCablecom architecture is to define a robust system that accommodates evolving technology without creating a legacy burden.

1.2.2 Purpose of this Recommendation

The purpose of this Recommendation is to specify profiles for codecs, packetization rules, encodings, and quality metrics to assure successful media interworking within an IPCablecom network and between an IPCablecom network and interconnecting networks including the PSTN and cellular networks. The media specified in this Recommendation include narrow-band audio, wideband audio, image, video and real-time text for communications services.

The actual codecs and other requirements that are mandatory, recommended, or optional depend on the functional component within the IPCablecom architecture, and the intended capability of the device or application. These requirements are specified in separate application capability documents.

This Recommendation is issued to facilitate component design and qualification testing leading to the manufacturability and interoperability of conforming hardware and software by multiple vendors.

1.2.3 High-level requirements for IPCablecom2

IPCablecom2 media stream transport and encoding design goals include the following:

- minimize the effects of latency, packet loss and jitter on sensitive media streams (e.g., voice and video) to ensure quality in the target environments (including audio/video telephony, IP video streaming and wireless);
- define a set of audio and video codecs and associated media transmission protocols that may be supported;
- accommodate emerging narrow-band and wideband voice codec technologies;
- accommodate emerging video codec technologies to provide support for applications such as video telephony, IP video streaming, etc.;
- accommodate real-time text conversation sessions;
- specify minimum requirements for echo cancellation and voice activity detection;
- support transparent, error-free dual tone multi frequency (DTMF) transmission;
- support for fax relay, modem relay, DTMF relay, and teletype (TTY);
- support calculation and reporting of voice quality metrics.

2 References

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

[ITU-T F.700] ITU-T Recommendation F.700 (2000), *Framework Recommendation for multimedia services*.

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

[ITU-T F.724] ITU-T Recommendation F.724 (2005), *Service description and requirements for videotelephony services over IP networks*.

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

[ITU-T G.114] ITU-T Recommendation G.114 (2003), *One-way transmission time*.

[ITU-T G.168] ITU-T Recommendation G.168 (2004), *Digital network echo cancellers*.

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

[ITU-T G.711 App.I] ITU-T Recommendation G.711 Appendix I (1999), *A high quality low-complexity algorithm for packet loss concealment with G.711*.

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

[ITU-T G.712] ITU-T Recommendation G.712 (2001), *Transmission performance characteristics of pulse code modulation channels*.

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

[ITU-T G.722.2] ITU-T Recommendation G.722.2 (2003), *Wideband coding of speech at around 16 kbit/s using Adaptive Multi-Rate Wideband (AMR-WB)*.

[ITU-T H.262] ITU-T Recommendation H.262 (2000) | ISO/IEC 13818-2:2000, *Information technology – Generic coding of moving pictures and associated information: Video*.

[ITU-T T.38] ITU-T Recommendation T.38 (2005), *Procedures for real-time Group 3 facsimile communication over IP networks*.

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

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

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

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

[ANSI T1.508] ANSI T1.508 (2003), *Loss Plan for Digital Networks*.

[ANSI/TIA-127-A] ANSI/TIA 127-A (2004), *Enhanced Variable Rate Codec Speech Service Option 3 for Wideband Spread Spectrum Digital Systems*.

[ATSC A/52B] ATSC A/52B, *Digital Audio Compression (AC-3, E-AC-3), Standard*, Rev. B, 14 June 2005.

[ETSI TS 102 005] ETSI TS 102 005 V1.3.1 (2007), *Digital Video Broadcasting (DVB); Specification for the use of Video and Audio Coding in DVB services delivered directly over IP protocols*.

[ETSI TS 126 091] ETSI TS 126 091 V6.0.0 (2004), *AMR speech codec; Error concealment of lost frames*.

[ETSI TS 126 092] ETSI TS 126 092 V6.0.0 (2004), *AMR speech codec; Comfort noise for AMR speech traffic channels*.

[ETSI TS 126 093] ETSI TS 126 093 V6.1.0 (2006), *AMR speech codec; Source controlled rate operation*.

[ETSI TS 126 094] ETSI TS 126 094 V6.1.0 (2006), *AMR speech codec; Voice activity detector for AMR speech traffic channels*.

[ETSI TS 126 101] ETSI TS 126 101 V6.0.0 (2005), *Adaptive Multi-Rate* (*AMR) speech codec frame structure*.

[ETSI TS 126 140] ETSI TS 126 140 V7.1.0 (2007), *Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications Systems (UMTS); Multimedia Messaging Service (MMS); Media formats and codes*.

[ETSI TS 126 190] ETSI TS 126 190 V6.1.1 (2005), *Adaptive Multi-Rate – Wideband (AMR-WB) speech codec; Transcoding functions*.

[ETSI TS 126 191] ETSI TS 126 191 V6.0.0 (2004), *AMR speech codec, wideband; Error concealment of lost frames*.

[ETSI TS 126 192] ETSI TS 126 192 V6.0.0 (2004), *Mandatory speech codec speech processing functions AMR Wideband speech codec; Comfort noise aspects*.

[ETSI TS 126 193] ETSI TS 126 193 V6.1.0 (2006), *Adaptive Multi-Rate – Wideband (AMR-WB) speech codec; Source controlled rate operation*.

[ETSI TS 126 194] ETSI TS 126 194 V6.0.0 (2004), *Mandatory speech codec speech processing functions AMR Wideband (AMR-WB) speech codec; Voice activity detector (VAD).*

[ETSI TS 126 201] ETSI TS 126 201 V6.0.0 (2004), *AMR speech codec,wideband; Frame structure*.

[ETSI TS 126 234] ETSI TS 126 234 V.6.10.0 (2006), *Transparent end-to-end packet-switched streaming service (PSS); Protocols and codecs*.

[ETSI TS 126 235] ETSI TS 126 235 V6.4.0 (2005), *Packet switched conversational multimedia applications; Default codecs*.

[ETSI TS 126 236] ETSI TS 126 236 V6.4.0 (2005), *Packet switched conversational multimedia applications; Transport protocols*.

[ETSI TS 126 290] ETSI TS 126 290 V6.3.0 (2005), *Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications Systems (UMTS); Audio codec processing functions; Extended Adaptive Multi-Rate – Wideband (AMR-WB+) Codec; Transcoding functions*.

[ETSI TS 126 346] ETSI TS 126 346 V7.6.0 (2008), *Universal Mobile Telecommunications Systems (UMTS); Multimedia Broadcast/Multicast Service (MBMS); Protocols and codecs*.

[ETSI TS 126 401] ETSI TS 126 401 V6.2.0 (2005), *Universal Mobile Telecommunications Systems (UMTS); General audio codec audio processing functions; Enhanced aacPlus general audio codec, General description (Release 6)*.

[ETSI TS 126 402] ETSI TS 126 402 V6.1.0 (2005), *Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications Systems (UMTS); General audio codec audio processing functions; Enhanced aacPlus general audio codec, Additional decoder tools (Release 6)*.

[ETSI TS 126 403] ETSI TS 126 403 V.6.0.0 (2005), *Universal Mobile Telecommunications Systems (UMTS); General audio codec audio processing functions; Enhanced aacPlus general audio codec, Encoder specification AAC part (Release 6)*.

[ETSI TS 126 404] ETSI TS 126 404 V.6.0.0 (2005), *Universal Mobile Telecommunications Systems (UMTS); General audio codec audio processing functions; Enhanced aacPlus general audio codec, Encoder specification SBR part (Release 6)*.

[ETSI TS 126 405] ETSI TS 126 405 V6.1.0 (2005), *Universal Mobile Telecommunications Systems (UMTS); General audio codec audio processing functions; Enhanced aacPlus general audio codec, Encoder specification parametric stereo part (Release 6)*.

[ETSI TS 126 410] ETSI TS 126 410 V6.9.0 (2007), *Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications Systems (UMTS); General audio codec audio processing functions; Enhanced aacPlus general audio codec, Floating-point ANSI C-code (Release 6)*.

[ETSI TS 126 411] ETSI TS 126 410 V6.6.0 (2007), *Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications Systems (UMTS); General audio codec audio processing functions; Enhanced aacPlus general audio codec, Fixed-point ANSI C-code (Release 6)*.

[ID SIP RTCP] *SIP Service Quality Reporting Event, draft-johnston-sipping-rtcp-summary-08, Alan Johnston et al*, October 14, 2005.

[IETF RFC 2198] IETF RFC 2198 (1997), *RTP Payload for Redundant Audio Data*.

[IETF RFC 2250] IETF RFC 2250 (1998), *RTP Payload Format for MPEG1/MPEG2 Video*.

[IETF RFC 2429] IETF RFC 2429 (1998), *RTP Payload Format for the 1998 Version of ITU-T Rec. H.263 Video (H.263+)*.

[IETF RFC 2733] IETF RFC 2733 (1999), *An RTP Payload Format for Generic Forward Error Correction*.

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

[IETF RFC 3016] IETF RFC 3016 (2000), *RTP Payload Format for MPEG-4 Audio/Visual Streams*.

[IETF RFC 3264] IETF RFC 3264 (2002), *An Offer/Answer Model with Session Description Protocol (SDP)*.

[IETF RFC 3267] IETF RFC 3267 (2002), *Real-Time Transport Protocol (RTP) Payload Format and File Storage Format for the Adaptive Multi-Rate (AMR) and Adaptive Multi-Rate Wideband (AMR-WB) Audio Codecs*.

[IETF RFC 3407] IETF RFC 3407 (2002), *Session Description Protocol (SDP) Simple Capability Declaration*.

[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 3555] IETF RFC 3555 (2003), *MIME Type Registration of RTP Payload Formats*.

[IETF RFC 3558] IETF RFC 3558 (2003), *RTP Payload Format for Enhanced Variable Rate Codecs (EVRC) and Selectable Mode Vocoders (SMV)*.

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

[IETF RFC 3640] IETF RFC 3640 (2003), *RTP Payload Format for Transport of MPEG-4 Elementary Streams*.

[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 3984] IETF RFC 3984 (2005), *RTP Payload Format for H.264 Video*.

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

[IETF RFC 4184] IETF RFC 4184 (2005), *RTP Payload Format for AC-3 Audio*.

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

[IETF RFC 4348] IETF RFC 4348 (2006), *Real-Time Transport Protocol (RTP) Payload Format for the Variable-Rate Multimode Wideband (VMR-WB) Audio Codec*.

[IETF RFC 4352] IETF RFC 4352 (2006), *RTP Payload Format for the Extended Adaptive Multi-Rate Wideband (AMR-WB+) Audio Codec*.

[IETF RFC 4424] IETF RFC 4424 (2006), *Real-Time Transport Protocol (RTP) Payload Format for the Variable-Rate Multimode Wideband (VMR-WB) Extension Audio Codec*.

[IETF RFC 4566] IETF RFC 4566 (2006), *SDP: Session Description Protocol*.

[IETF RFC 4598] IETF RFC 4598 (2006), *Real-time Transport Protocol (RTP) Payload Format for Enhanced AC-3 (E-AC-3) Audio*.

[ISO/IEC 11172-3] ISO/IEC 11172-3:1993, *Information technology – Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s – Part 3: Audio*.

[ISO/IEC 13818-3] ISO/IEC 13818-3:2007, *Information technology – Generic coding of moving pictures and associated audio information – Part 3: Audio*.

[ISO/IEC 13818-7] ISO/IEC 13818-7:2006, *Information technology – Generic coding of moving pictures and associated audio information – Part 7: Advanced Audio Coding (AAC). (MPEG2)*

[ISO/IEC 14496-1] ISO/IEC 14496-1:2004, *Information technology – Coding of audio-visual objects – Part 1: Systems. (MPEG4)*

[ISO/IEC 14496-2] ISO/IEC 14496-2: 2004, *MPEG-4, Coding of audio-visual objects – Part 2: Visual*.

[ISO/IEC 14496-3] ISO/IEC 14496-3:2005, *Information technology – Coding of audio-visual objects – Part 3: Audio. (MPEG-4)*

[TIA/IS-893] TIA/IS-893 (2005), *Selectable Mode Vocoder (SMV) Service Option for Wideband Spread Spectrum Communication Systems*.

[TIA-1016-A] TIA-1016-A (2006), *Source-Controlled Variable-Rate Multimode Wideband Speech Codec (VMR-WB), Service Options 62 and 63 for Spread Spectrum Systems*.

3 Terms and definitions

This Recommendation uses the following terms:

3.1 active: A service flow is said to be "active" when it is permitted to forward data packets. A service flow must first be admitted before it is active.

3.2 authentication: The act of giving access to a service or device if one has permission to have the access.

3.3 downstream: The direction from the headend toward the subscriber location.

3.4 dynamic quality of service: A quality of service assigned on the fly for each communication depending on the QoS requested.

3.5 encryption: A method used to translate plaintext into ciphertext.

3.6 endpoint: A terminal, gateway or multipoint conference unit (MCU).

3.7 gateway: Devices bridging between the IPCablecom IP voice communication world and the PSTN. Examples are the media gateway, which provides the bearer circuit interfaces to the PSTN and transcodes the media stream, and the signalling gateway, which sends and receives circuit switched network signalling to the edge of the IPCablecom network.

3.8 header: Protocol control information located at the beginning of a protocol data unit.

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

3.10 key: A mathematical value input into the selected cryptographic algorithm.

3.11 latency: The time taken for a signal to pass through a device or network.

3.12 media gateway: Provides the bearer circuit interfaces to the PSTN and transcodes the media stream.

3.13 network management: The functions related to the management of data across the network.

3.14 off-net call: A communication connecting an IPCablecom subscriber out to a user on other networks such as PSTN or cellular.

3.15 one-way hash: A hash function that has an insignificant number of collisions upon output.

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

3.17 privacy: Also known as confidentiality. A way to ensure that information is not disclosed to anyone other than the intended parties. Information is usually encrypted to provide confidentiality.

3.18 proxy: A facility that indirectly provides some service or acts as a representative in delivering information, thereby eliminating the need for a host to support the service.

3.19 pulse code modulation: A common method of digitizing an analogue signal (such as a human voice) into a bit stream using simple analogue to digital conversion techniques. [ITU-T G.711] defines its use in the PSTN with two encoding laws, μ-law, used in North America, and A-law, used elsewhere.

3.20 quality of service: Guarantees network bandwidth and availability for applications.

3.21 real-time text conversation: Text conversation sessions as specified in [ITU-T T.140] and [IETF RFC 4103].

3.22 real-time transport protocol: A protocol for encapsulating encoded voice and video streams. Refer to [IETF RFC 3550].

3.23 session initiation protocol: An application-layer control (signalling) protocol for creating, modifying, and terminating sessions with one or more participants.

3.24 session initiation protocol plus: An extension to SIP.

3.25 terminal adapter: A device that converts an analogue tip and ring interface into a digital signal; it includes a hybrid to convert the interface from 2-wire to 4-wire.

3.26 upstream: The direction from the subscriber location toward the headend.

3.27 user datagram protocol: A connectionless protocol built upon Internet Protocol (IP).

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

ASO	Arbitrary Slice Ordering
AVC	Advanced Video Coding
CABAC	Context-based Adaptive Binary Arithmetic Coding
CAVLC	Context-Based Adaptive Variable Length Coding
CIF	Common Intermediate Format
CNG	Comfort Noise Generation
Codec	COder-DECoder
CPE	Customer Premises Equipment
CSRC	Contributing Source
DQoS	Dynamic Quality of Service
DTMF	Dual Tone Multi Frequency (tones)
DTX	Discontinuous Transmission
FMO	Flexible Macroblock Ordering
HD	High Definition
HFC	Hybrid Fibre/Coax
IP	Internet Protocol
MG	Media Gateway
MGCP	Media Gateway Control Protocol (refer to [b-IETF RFC 3435])

MMD	Multimedia Domain
NAT	(IP) Network Address Translation
NCS	Network Call Signalling
PCM	Pulse Code Modulation
PLC	Packet Loss Concealment
PSTN	Public Switched Telephone Network
QCIF	Quarter Common Intermediate Format
QoS	Quality of Service
RDA	Rate-Determination Algorithm
RFC	Request for Comments
RSVP	Resource reSerVation Protocol
RTCP	Real-Time Control Protocol
RTP	Real-Time Transport Protocol
SD	Standard Definition
SDP	Session Description Protocol
SID	Silence Insertion Descriptor
SIP	Session Initiation Protocol
SIP+	Session Initiation Protocol Plus
SSRC	Synchronizing Source
TDM	Time Division Multiplexing
UDP	User Datagram Protocol
VAD	Voice Activity Detection
VBR	Variable Bit-Rate
VoIP	Voice over IP

5 Conventions

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

"MUST"	This word 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 means that this item is truly optional. For example, one vendor may choose to include the item because a particular marketplace requires it or because it enhances the product; another vendor may omit the same item.

6 Common criteria for media transport over IP

This clause outlines the required basic functionality of the IPCablecom2 architecture with respect to audio, video and real-time text media handling. The key requirement for narrow-band voice communications using IP transmission over a cable infrastructure is the ability to attain "toll" or better audio quality. This applies to both on-net calls and off-net calls to other networks such as the PSTN or cellular. The toll quality standard is also the objective to be met on calls to cellular networks using the latest generation of codecs, recognizing that this requires optimum radio conditions. IPCablecom2 additionally provides the means for cable operators to offer superior audio communications quality, exceeding current PSTN standards, by using wideband audio codecs. Finally, IPCablecom2 provides the capability of two-way and multi-way video communications. Given the variable nature of shared packet mediums and the stringent human-factor requirements of perceived communications quality, it is necessary to optimize multiple system parameters to attain quality goals at reasonable cost.

6.1 IP network criteria for codec support

6.1.1 Packet loss control

There is a direct correlation between packet loss and audio quality. For voice communications, this effect can be masked by packet loss concealment techniques up to approximately 2-3% packet loss rate. Above this loss rate, speech quality degrades rapidly even with packet loss concealment. However, packet loss concealment has no benefit for voiceband data and without introducing packet redundancy, voiceband data transmissions require packet loss rates of 10^{-5} or better to avoid call failures. Similarly, most video codecs of interest rely on inter-frame compression, thus are highly sensitive to packet loss, especially of key frames. Applications and codecs more sensitive to packet loss may provide redundancy or error correction, which increases latency through buffering.

6.1.2 Latency control

Standards for network latency are defined in [ITU-T G.114]. Specifically, an end-to-end delay of 150 ms is considered acceptable for most interactive applications. Achieving this requires a coordinated system design in terms of the application and the system resources.

There are multiple device elements and network components inducing latency during traversal of an audio or video signal. The primary contributors to latency are as follows:

- Audio or video sampling and analogue-to-digital conversion.
- Buffering of samples (framing, plus look-ahead).
- Compression processing.
- Packetization of compressed data.
- Access network traversal.
- Routing to the backbone network.
- Backbone traversal (propagation delay).
- Far-end reception of packets and traversal of local access network.
- Buffering of out-of-order and delayed packets.
- Decoding, decompression, and reconstruction of the media stream.

Some of the methods to control latency are described in the following clauses.

6.1.2.1 Jitter buffer management

Jitter buffers are required to smooth out packet delay variation to provide for a continuous play out on a time division multiplexing (TDM) or analogue interface. Setting the optimum size of the nominal jitter buffer is a compromise between latency and packet loss. Voice, being sensitive to latency but tolerant of a certain amount of packet loss, performs best with an adaptive jitter buffer which, in the absence of jitter, will reduce its nominal size to some preconfigured minimum. Voiceband data, however, while less sensitive to latency, is intolerant of packet loss and so performs best with a fixed jitter buffer that does not attempt to reduce during periods of low jitter. In the event of a packet arriving after its required play-out time, both types of jitter buffer will typically adjust their nominal play-out timing to match the late arrival. Later, an adaptive jitter buffer may adapt down by discarding packets instead of playing them out. A fixed jitter buffer will not adapt down.

Configurable jitter buffer parameters typically include the minimum and maximum values.

6.1.2.2 Framing and packetization

One way to minimize latency (and the effect of packet loss) is to send small packets containing the minimum number of frames. However, this may increase bandwidth use by increasing the header-to-data ratio for packets. This suggests that the optimal packet size for voice applications is fairly small, containing compressed information for 10, 20 or 30 ms of audio (typically one, two, or, at most, three frames of compressed audio data). These packet sizes are only applicable to audio data as the frame sizes for video data are variable.

To avoid additional buffering delay, packets are sent at a rate equal to integral multiples of the audio sample frame rate of the codec. This synchronization results in lock-step operation between the codec framing and packet generation.

6.1.2.3 Codec selection

The frame size is a direct contributor to delay. It is dependent on the codec.

6.1.3 Codec transcoding minimization

Given the rate of introduction of new network technologies and their associated codecs, it quickly becomes apparent that it is not cost-effective for every IPCablecom2 device to support every possible codec technology that could be interconnected with an IPCablecom system. Transcoding within the IPCablecom network is inevitable. Transcoding is often associated with undesirable artefacts such as degraded voice quality and increased latency. However, the use of a high-quality and low-delay codec mitigates degradation and delay.

6.1.4 Bandwidth minimization

There are three 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.
- Large packet sizes may be used, containing multiple audio frames.
- 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. In either case, selection takes place after the initial capabilities exchange to determine a compatible codec between endpoints, and assumes that the required bandwidth is available.

Variable rate transmission employs methods resulting in a non-constant rate media stream. For example, voice activity detection (VAD) with 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 bit stream to adapt the source data to network conditions as, for example, in cellular networks.

6.2 Overall quality targets

With respect to media quality, the IPCablecom2 architecture should be designed to meet the following end-to-end performance targets: ([b-DESK CONF], [ITU-T G.114] and [b-ETSI TR 102 493]):

- Mouth-to-ear delay (end-to-end) for audio telephony: ≤150 ms.
- Mouth-to-ear delay (end-to-end) for video telephony: ≤150 ms.
- End-to-end jitter for audio telephony: ≤50 ms.
- End-to-end jitter for video telephony: ≤130 ms.
- Packet loss (in the absence of packet loss concealment): ≤1 %.
- Skew between audio and associated video:
  - ≤20 ms audio advance over video.
  - ≤120 ms audio delay following video.
- Video frame rate: ≥24 frames per second.

NOTE – It is recognized that 150 ms end-to-end delay for video is unreasonable for many existing video conferencing systems. However, to achieve a consistent end-user experience it is recommended that the delay objective remain the same.

6.3 Media security

There are no security requirements defined for media transmission. For more details see Appendix III to [b-ITU-T J.360].

7 Gap analysis between 3GPP, 3GPP2 and IPCablecom codec requirements

This clause identifies the differences in capability between the audio and video codecs specified by 3GPP for use in IMS, and by 3GPP2 for use in MMD, and the requirements and associated features applicable to IPCablecom codecs. In addition to the codec applications within the IPCablecom VoIP network, which may include the use of dual-mode cellular-VoIP handsets, consideration is given to interworking with the legacy PSTN as well as with 3GPP and 3GPP2 cellular networks, including IMS.

7.1 Audio codecs

7.1.1 3GPP and 3GPP2 audio codecs

3GPP and 3GPP2 specify audio codecs and codec capabilities designed to optimize voice quality under variable radio conditions, while minimizing use of expensive spectrum capacity. Toll quality voice communications is provided under optimum radio conditions between cellular users where transcoding can be avoided and between cellular users and the PSTN where transcoding to and from G.711 is required. Under degraded radio conditions, 3GPP and 3GPP2 trade off voice quality in order to avoid drop-outs. This is done by increasing the error-correcting redundant coding while dropping the bandwidth available to the voice codec.

Recognizing that toll quality voice is not possible with more than one low bit-rate coding, a method has been standardized for transporting low bit-rate compressed speech across a G.711 tandem TDM between cellular networks. This is known as tandem-free operation (TFO) or transcoder-free operation. It is limited to operation between two 3GPP networks or two 3GPP2 networks.

Over the air interface, 3GPP and 3GPP2 specify an out-of-band signalling method for the transport of DTMF signals since the audio codec is incapable of accurately rendering the dual-tones. However, an in-band text transmission method using a relatively low-frequency modem is specified for support of TTY devices.

Over a VoIP network, 3GPP in [ETSI TS 126 234] and [b-3GPP TS 24.229] recommends the use of DTMF [IETF RFC 2833] telephone events to assure reliable transmission.

For data services, 3GPP and 3GPP2 specify a packet-based radio access network that provides for the transport of unmodulated data. Only at the interworking point to the PSTN would this data be modulated into audio. Other than to support TTY devices, there is no provision for the transport of modulated data within 3GPP and 3GPP2 voice codecs.

In addition to the traditional 3.1 kHz voice codecs, both 3GPP and 3GPP2 specify wideband audio codecs based on a 16 kHz sampling rate to allow for higher-quality voice communications than is possible in the PSTN. This provides cellular operators an opportunity for differentiation of their 3rd generation services over previous cellular services and, especially, over the PSTN which may thereby accelerate the substitution of the legacy wireline network.

Table 7-1 summarizes the audio codecs chosen for 3GPP and 3GPP2.

Table 7-1 – Summary of 3GPP and 3GPP2 audio codecs

	3GPP	3GPP2
Narrow-band telephony	AMR	SMV EVRC
Wideband telephony	AMR-WB	VMR-WB

7.1.2 Audio codec analysis for IPCablecom

One of the mandates of IPCablecom2 is to provide a set of audio codecs to serve in a variety of environments including, but not limited to, cellular, and to provide for interworking with E-MTAs. Hence, IPCablecom2 codec selection includes additional codecs beyond those specified by 3GPP and 3GPP2. IPCablecom also imposes strict requirements for the reliable transport of various types of voiceband data not encountered in cellular networks, and is therefore, beyond the scope of 3GPP or 3GPP2. Faxes, dial-up modems, point-of-sale terminals and telephone devices for the deaf (TDD) are all examples of equipment that use voiceband data that are reliably supported by the PSTN and must be well supported by IPCablecom networks.

7.2 Video codecs

7.2.1 3GPP and 3GPP2 video codecs

3GPP and 3GPP2 specify mandatory and recommended codecs to allow cellular video telephony and streaming with QCIF resolution (176 × 144), consistent with a cellular handset's small screen size and cellular networks' limited bandwidth [ETSI TS 126 234] and [ETSI TS 126 235]. These requirements are summarized in Table 7-2.

Table 7-2 – Summary of 3GPP and 3GPP2 video codecs

	3GPP IMS	3GPP2 MMD
Video telephony or entertainment (QCIF)	H.263 Profile 0 @ level 45 (mandatory) Profile 3 @ level 45 (recommended) H.264 (recommended) baseline profile @ level 1b, with constraint_set1_flag = 1 MPEG-4 part 2 (recommended) Simple profile @ level 0b	H.263 Profile 0 @ level 45 (mandatory) Profile 3 @ level 45 (recommended) H.264 (recommended) baseline profile @ level 1b, with constraint_set1_flag = 1 MPEG-4 part 2 (recommended) Simple profile @ level 0b

7.2.2 Video codec analysis for IPCablecom2

IPCablecom2 requires video codecs to be supported for various types of consumer devices, especially personal computers. 3GPP and 3GPP2 support well-known video codecs, but these codecs are limited to profiles for screen resolutions suitable for cellular devices. IPCablecom2 defines additional codecs and profiles to support multiple resolutions as is necessary to provide user satisfaction with both smaller-screen and larger-screen devices.

7.3 Quality metrics

7.3.1 3GPP and 3GPP2 VoIP quality metrics

No VoIP or other IP quality metrics have been defined by 3GPP for use in IMS or by 3GPP2 for use in MMD. Both systems rely on the existing local RTP loss, jitter, and round-trip delay performance measurements.

7.3.2 Quality metric analysis for IPCablecom

IPCablecom networks need to be able to identify and, where possible, locate problems for both on-net sessions and when interworking with other networks including the PSTN and cellular networks. IPCablecom improved upon the basic local RTP statistics, first, through the provision of remote statistics and, second, through support of the RTCP XR VoIP metrics package specified in [IETF RFC 3611].

Although the IMS and 3GPP codec specifications for conversational multimedia presently do not specify use of RTCP XR quality metrics, it is important to partition problems between IPCablecom networks and peer IMS networks, including 3GPP or 3GPP2 cellular networks. Additionally, applications based on IPCablecom architecture may utilize VoIP metrics to resolve voice quality issues. Therefore, the support of the VoIP metrics block for conversational applications is defined for IPCablecom UE.

8 Codec and media requirements

This clause specifies codec and RTP media requirements.

Figure 8-1 below, shows an abstract architecture for the media flows in an IPCablecom network. Originating or terminating media at the customer premises is the user equipment (UE) or client device which may be a dedicated VoIP phone or videoconferencing unit, a soft VoIP phone or videoconference terminal running on a PC, or an embedded media terminal adapter that provides POTS interfaces to legacy telephone, fax, modem, or TTY terminals. Processing media within the IPCablecom network are one of three types of device depending on the destination of the VoIP session. For VoIP calls between client devices and the PSTN, a media gateway (MG) terminates the VoIP session at the edge of the IPCablecom network and provides TDM trunk interfaces to the PSTN.

VoIP and video sessions between client devices in the IPCablecom network and another VoIP network must traverse a media proxy, which acts as a gateway to protect the IPCablecom network through network address translation (NAT), firewall, and other security and flow policing functions. The media proxy may transcode between different media formats when necessary. The media proxy may also perform a role in VoIP performance monitoring, in particular for problem sectionalization between the IPCablecom network and a peer VoIP network.

VoIP or video sessions from client devices may also terminate at a media server within the IPCablecom network. For audio, a media server may be an announcement source, a recorder, an interactive function such as voice mail, or a conference bridge. Similarly, media servers for video may originate media, record or otherwise process video, or bridge video for live conferences.

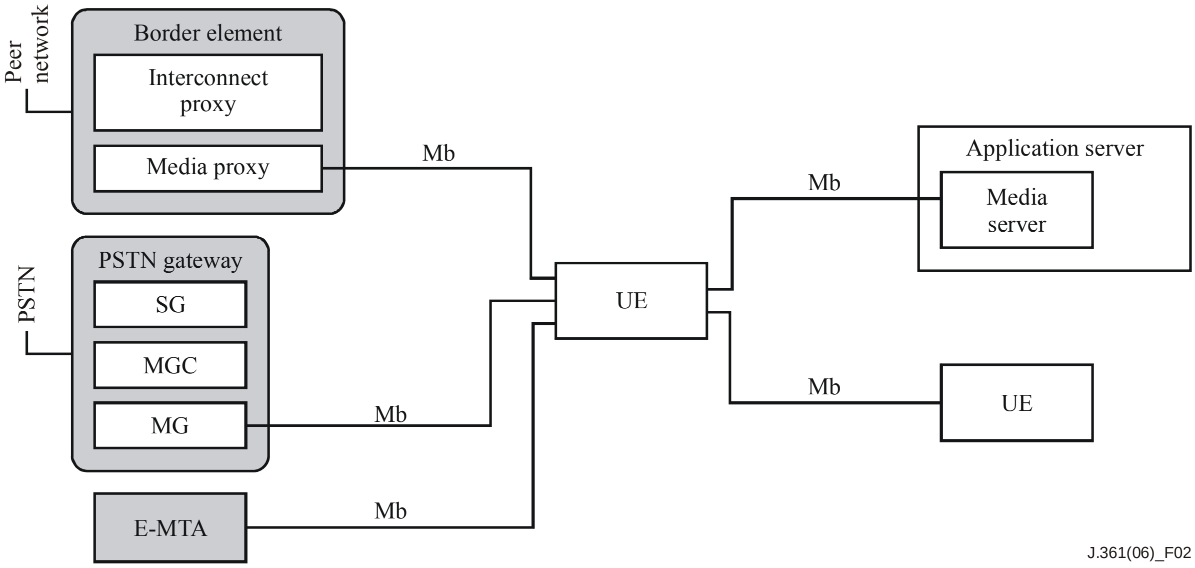

Figure 8-1 – RTP/RTCP media connection

The interfaces depicted in Figure 8-1 are described in Table 8-1.

Table 8-1 – Media stream interfaces

Reference point	IPCablecom network elements	Reference point description
Mb	UE – UE UE – MG UE – Border element UE – AS (media server) UE – E-MTA	Allows media-capable components to send and receive media data packets. Specifically, a UE can exchange media with another UE, a MG, an application server, border element and an E-MTA.

The media travelling across the Mb reference point can be audio traffic encoded by narrow-band or wideband audio codecs, fax image traffic, and video traffic encoded by video codecs, or the combination of audio and video traffic types.

In this abstract representation of media flows, each endpoint contains one or more codecs to allow audio or video communication between endpoints.

Not depicted in Figure 8-1, but shown in Figure 1-1, is the TURN server that provides data relay functions in support of NAT traversal for media flows in and out of the CMTS.

Codec requirements in this clause are specified with reference to user equipment (UE) and media gateways (MG). Unless there is text to the contrary, the term media gateway should be interpreted to encompass the media proxy and media server functions in addition to the PSTN media gateway functions.

8.1 RTP requirements

User equipment and media gateways MUST support the real-time transport protocol (RTP), as defined in [IETF RFC 3550] and [IETF RFC 3551], for transport of audio and video media flows.

A media server that performs mixing of RTP streams MAY transmit contributing source (CSRC) lists. 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.

To facilitate traversal of NAT and firewall gateways, user equipment and media gateways MUST transmit their RTP stream from the same IP address and port in which it has advertised to receive RTP on in its SDP. Similarly, user equipment and media gateways MUST transmit their RTCP stream from the same IP address port in which it has advertised to receive RTCP on in its SDP description via the a=rtcp attribute.

8.2 RTCP requirements

To facilitate vendor interoperability, the following RTCP profile has been defined for user equipment and media gateways. In the event that a discrepancy arises between the RFCs and this profile, this profile takes precedence.

8.2.1 General requirements of the IPCablecom RTCP profile

User equipment and media gateways MUST send and receive RTCP messages as described in [IETF RFC 3550] and [IETF RFC 3551] subject to the following over-riding requirements and clarifications.

User equipment and media gateways 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 to be 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, user equipment or media gateways MAY stop sending RTCP packets to a remote endpoint if an ICMP port unreachable error 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, user equipment or media gateways MAY stop sending RTCP packets to a remote endpoint if no RTCP packets have been received within five report transmission intervals. This requirement allows the endpoint to stop sending RTCP packets to user equipment or media gateways that simply receive and discard RTCP reports.

User equipment and media gateways SHOULD provide a configurable RTCP transmission interval with a default average of 5 s. The transmission interval chosen MUST be randomized over the range of 0.5 of the average to 1.5 times the average as described in [IETF RFC 3550]. For multi-party conference calls without the use of a bridge, user equipment and media gateways MAY support the RTCP interval calculation method described in clauses 6.2 and 6.3 of [IETF RFC 3550].

User equipment and media gateways MUST receive RTCP messages sent by the remote communication peers. User equipment and media gateways MUST NOT require RTCP message for operation. 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 hybrid fibre/coax (HFC) access network bandwidth utilization, admission control, call signalling, and DOCSIS signalling, and remains for further study.

Synchronizing source (SSRC) collision detection and resolution is optional for user equipment and media gateways that are capable of unambiguously distinguishing between media packets and reports that they send and those that they receive. 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 user equipment and media gateways that are establishing unicast, point-to-point connections carrying one RTP stream. If SSRC collision detection and resolution is supported, one or both of the user equipment or media gateways MUST resolve SSRC collisions as follows:

1) send BYE;
2) select new SSRC; and
3) send sender description with new SSRC.

SSRC collision detection and resolution is OPTIONAL for user equipment and media gateways 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 media 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; and
3) send sender description with new SSRC.

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

source description (SDES): 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 but, 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.

sender report (SR): MUST be sent by user equipment and media gateways transmitting RTP packets (as described in [IETF RFC 3550]) except, as previously described, when errors occur or the remote endpoint does not send RTCP packets, in which case they MAY be sent.

receiver report (RR): 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.

application-defined (APP): MAY be sent as implementation needs dictate and MUST NOT contain identity information. User equipment and media gateways MUST ignore and silently discard APP messages with unrecognized contents.

goodbye (BYE): MUST be sent upon RTP connection deletion or when renegotiating SSRC upon collision detection and resolution (see below). User equipment and media gateways MUST send BYE commands when the application needs to discontinue use of an SSRC and start a new SSRC, for example, on media gateway failover. User equipment and media gateways MUST NOT use BYE messages to indicate or detect any call progress condition. For example, user equipment or media gateways 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 signalling protocol, such as SIP.

NOTE – 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.

8.2.2 Standard statistics reporting

RTCP sender reports and receiver reports include a reception report block to transfer basic loss and jitter measurements. Loss information is transferred as both a cumulative session count of lost packets and a loss rate corresponding with the RTCP reporting period. The SR/RR reception report block also includes timestamps to enable a near-end round-trip delay calculation to up to 1/65'536 second accuracy.

User equipment and media gateways MUST include loss and jitter measurements in transmitted sender or receiver reports. User equipment and media gateways must store loss and jitter metrics received in arriving sender or receiver reports until the next RTCP SR or RR arrives for the same session. User equipment and media gateways MUST perform round-trip delay calculations based on the exchange of RTCP SR/RR with the far end media endpoint.

8.2.3 Extended statistics reporting using RTCP XR

The RTCP extended reports (XR) as defined in [IETF RFC 3611] MAY be sent by user equipment and media gateways as appropriate for the type of media, and if negotiated, on a given connection. IPCablecom presently only defines use of the VoIP metrics report block as described in clause 8.7.1, but user equipment and media gateways MAY send other RTCP XR payload types. User equipment and media gateways that are capable of sending RTCP XR reports MUST be capable of receiving, interpreting and parsing the corresponding RTCP XR report blocks.

8.3 General session description for 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 [b-IETF RFC 3261]. This clause describes the required specification of the codec in SDP, and the required mapping of the SDP description into 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 4566] are dynamically bound by using a dynamic payload type from the range 96-127, as defined in [IETF RFC 4566] and a media attribute line. For example, a typical SDP message for AMR would be composed as follows:

```
m=audio 3456 RTP/AVP 96
a=rtpmap:96 AMR/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 "AMR" with a clock rate of 8000 samples/s.

A typical example for H.263 video would be:

```
m=video 49170 RTP/AVP 98
a=rtpmap:98 H263-2000/90000
a=fmtp:98 profile=0; level=40
b=TIAS:2048000
```

8.3.1 SDP use

8.3.1.1 Attributes (a=)

```
a=<attribute> : <value>
a=fmtp:<format> <format specific parameters>

a=sqn: <sequence number>
a=cdsc: <capability number> <media> <transport> <media format list>
a=cpar: <capability parameter>
a=cparmin: <capability parameter>
a=cparmax: <capability parameter>

a=mptime: <list of packet times separated by space>
a=ptime: <packet time>
```

Send: One or more of the "a" attribute lines specified below MAY be included.

Receive: One or more of the "a" attribute lines specified below MAY be included and MUST be acted upon accordingly. Attribute values are case-insensitive. Implementations MUST accept the lower case, upper case, and mixed upper/lower case encodings of all attributes.

NOTE 1 – SDP [IETF RFC 4566] requires unknown attributes to be ignored.

`fmtp:`

Send: This field MAY be used to provide parameters specific to a particular format. For example, the field could be used to describe telephone events supported for an [IETF RFC 2833] format. When used, the format MUST be one of the formats specified for the media. The parameters specified are provided in a separate specification that details the usage of the format.

Receive: When used, the field MUST be used in accordance with [IETF RFC 4566].

NOTE 2 – Refer to clauses 8.3.2, 8.4.2.4.1 and 8.4.2.6.2 for more specific information regarding the "fmtp" attribute.

```
sqn:
cdsc:
cpar:
cparmin:
cparmax:
```

As defined in [IETF RFC 3407], together, these attributes form a capability set which describes the complete media capabilities of an endpoint. The capability set is declarative and the answer is independent of the offer.

Send: Offers and answers MAY include a capability set consistent with [IETF RFC 3407].

Receive: An offerer and answerer MAY interpret the capability set in an answer and offer, respectively.

Consider the following answer:

```
v=0
o=- 25678 753849 IN IP4 128.96.41.1
s=
c=IN IP4 128.96.41.1
t=0 0
a=pmft:T38
m=audio 3456 RTP/AVP 18 96
a=rtpmap:96 PCMU/8000
a=gpmd: 96 vbd=yes
a=sqn: 0
a=cdsc: 1 audio RTP/AVP 18 96 97
a=cpar: a=rtpmap:97 t38/8000
a=cdsc: 4 udptl t38
m=image 0 udptl t38
m=image 0 tcp t38
```

The "a=pmft:T38" SDP attribute indicates that T.38 is the preferred fax handling method (over V.152), yet the media descriptions suggest that T.38 is not supported at the time the SDP was exchanged (i.e., the endpoint does not support the T.38 autonomous transitioning method). However, the capability set explicitly indicates that the endpoint supports T.38 over UDPTL and T.38 over RTP (but not T.38 over TCP) as latent capabilities.

`mptime:`

This attribute is a media-level attribute defined by IPCablecom. The "mptime" attribute defines a list of packetization period values that the endpoint is capable of using (sending and receiving) for this connection.

Send: The "mptime" attribute MAY be present. If used, there MUST be precisely one entry in the list for each <format> entry provided in the "m=" line. Entry number j in this list defines the packetization period for entry number j in the "m=" line. The first entry in the list MUST be a decimal number whereas subsequent entries in the list MUST be either a decimal number or a hyphen ("-"). For those media formats where a single packetization rate does not apply (e.g., non-voice codecs such as telephone-event or comfort noise), a hyphen MUST be encoded at the corresponding location in the list of packetization periods.

Receive: Conveys the list of packetization periods that the remote endpoint is capable of using for this connection; one for each media format in the "m=" line. For media formats with packetization period specified as a hyphen ("-"), the endpoint MUST use one of the packetization periods that is actually specified in the list. If the "mptime" attribute is absent, then the value of the "ptime" attribute, if present, MUST be taken as indicating the packetization period for all codecs present in the "m=" line.

`ptime:`

Send: The "ptime" attribute MAY be sent in an offer. If the "ptime" attribute is received in an offer, it SHOULD be sent as part of the answer.

Receive: The field MUST be ignored if the SDP contains the "mptime" attribute, as required in IPCablecom-compliant devices. If the "mptime" attribute is not present, then this field is used to define the packetization interval for all codecs present in the SDP description.

NOTE 3 – [IETF RFC 4566] defines the "maxptime" SDP attribute and [ITU-T V.152] defines the "maxmptime" SDP attribute. The precedence of these attributes with respect to the "ptime" and "mptime" attributes is not presently defined.

8.3.2 Session description for audio/RED

The following SDP attributes are applicable to audio service use for [IETF RFC 2198].

```
a=<attribute> : <value>

a=rtpmap:<format> <encoding name>/<clock rate>[/<encoding parameters>]
a=rtpmap:<format> RED/8000

a=fmtp:<format> <format specific parameters>
a=fmtp:<format> <value>/<value>/…/<parameter>
a=fmtp:<format> 97/97
```

Send: One or more of the "a" attribute lines specified below MAY be included.

Receive: One or more of the "a" attribute lines specified below MAY be included and MUST be acted upon accordingly. Attribute values are case-insensitive. Implementations MUST accept the lower case, upper case and mixed upper/lower case encodings of all attributes.

NOTE – SDP [IETF RFC 4566] requires unknown attributes to be ignored.

`rtpmap:`

Send: When transmitting an offer, if [IETF RFC 2198] redundancy is supported and desired to be used, then the "rtpmap" attribute with the "RED" encoding name MUST be included. When transmitting an answer, if [IETF RFC 2198] redundancy is supported and desired to be used, and the offer included the "rtpmap" attribute with the "RED" encoding name, then the "rtpmap" attribute with the "RED" encoding name MUST be included and [IETF RFC 2198] redundancy MUST then be used. In all other cases, the "rtpmap" attribute with the "RED" encoding name MUST NOT be included and [IETF RFC 2198] MUST NOT be used.

Receive: [IETF RFC 2198] redundancy MUST NOT be used if the "rtpmap" attribute with the "RED" encoding name is absent.

`fmtp:`

Send: This attribute MUST be included as a recommendation in terms of the number of redundancy levels (primary, secondary, tertiary, …) and the media format associated with each level.

Receive: An offerer and answerer SHOULD honour the recommendation in the answer and offer, respectively.

The following is an example of the media representation in SDP for describing an [IETF RFC 2198] primary and secondary encoding (one level of redundancy) involving [b-ITU-T G.729]:

```
m=audio 49130 RTP/AVP 18 96
a=rtpmap: 96 RED/8000
a=fmtp: 96 18/18
```

8.4 Narrow-band codec specifications

Narrow-band codecs are defined to operate on audio signals bandpass filtered to a frequency range of 300-3400 Hz [ITU-T G.712] and sampled at 8000 samples/s. For codecs other than [ITU-T G.711], the input to the codec is generally in the form of 16-bit uniformly quantized samples with at least 13 bits of dynamic range. For [ITU-T G.711], the codec input is specified as 13-bit or 14-bit uniform PCM samples according to [ITU-T G.711]. A comparison of the well-known narrow-band codecs is provided in Appendix I.

8.4.1 Supported narrow-band codecs

The following clauses describe every narrow-band codec supported in IPCablecom. Whether a particular narrow-band codec is mandatory, recommended or optional depends on the application for which it is used. Therefore, the normative status of each codec is indicated in the associated application capability documents. However, if a particular codec is supported for an application, all the requirements for that codec as specified in this clause MUST be met.

8.4.1.1 G.711 codec

G.711 is the standard used by the PSTN to represent pulse code modulation (PCM) samples of signals of voice frequencies, sampled at the rate of 8000 samples/s. A G.711 encoder will create a 64 kbit/s bit stream. The standard has two forms of logarithmic quantization, viz., A-Law and μ-Law. An A-Law G.711 PCM encoder converts 13-bit linear PCM samples into 8-bit compressed PCM (logarithmic form) samples. A μ-Law G.711 PCM encoder converts 14-bit linear PCM samples into 8-bit compressed PCM samples. This codec provides toll quality voice and is ubiquitous in usage for narrow-band audio communications.

[ITU-T G.711] (both μ-law and A-law versions) MAY be supported by user equipment and media gateways. It can be used to provide "fallback" for services such as fax, modem, and services for the hearing-impaired, as well as common gateway transcoding support.

8.4.1.1.1 Packet loss concealment

For G.711, user equipment and media gateways SHOULD use the method defined in Appendix I of [ITU-T G.711].

8.4.1.1.2 Voice activity detection and silence suppression

G.711 does not have an associated VAD mechanism. User equipment and media gateways MAY employ VAD and silence suppression such as discontinuous transmission (DTX) to reduce bandwidth. If silence suppression is used, the user equipment and media gateways SHOULD transmit silence insertion descriptor frames as specified in Appendix II to [ITU-T G.711].

8.4.1.1.3 Payload header format

For G.711, no specific payload header format is specified. Standard RTP usage applies as per [IETF RFC 3550] and [IETF RFC 3551].

8.4.1.1.4 Session description

Parameters are mapped to SDP in a standard way. When conveying information by SDP, static payload types SHOULD be used, in accordance with [IETF RFC 3551].

The following is an example of the media representation in SDP for describing G.711 with μ-law PCM quantization and 10 ms packetization:

```
m=audio 49140 RTP/AVP 0
a=mptime:10
```

8.4.1.2 Internet low bit codec (iLBC)

iLBC was selected as a codec standard suitable for packet-based communication networks. Additionally, iLBC has undergone IETF standardization as part of the IETF audiovisual transport (AVT) working group [IETF RFC 3951], [IETF RFC 3952].

iLBC MAY be supported by user equipment and media gateways. iLBC provides two modes with coding rates of 13.3 kbit/s and 15.2 kbit/s using 30 ms and 20 ms frame sizes respectively. User equipment and media gateways that implement iLBC MUST support both modes of operation.

8.4.1.2.1 Packet loss concealment

For iLBC, user equipment and media gateways SHOULD use the method defined in [IETF RFC 3951] for packet loss concealment.

8.4.1.2.2 Voice activity detection and silence suppression

iLBC does not have an associated VAD mechanism. User equipment and media gateways MAY employ VAD and silence suppression such as discontinuous transmission (DTX) to reduce bandwidth. If silence suppression is used with iLBC then user equipment and media gateways SHOULD transmit silence insertion descriptor frames as specified in Appendix II to [ITU-T G.711].

8.4.1.2.3 Payload header format

User equipment and media gateways MUST support the payload header format as specified in [IETF RFC 3952] for iLBC. A standard RTP header is used along with one or more frames of iLBC to form the packet. User equipment and media gateways MUST use the codec payload bit packing as specified in [IETF RFC 3952] for iLBC. There are no options specific to this payload header format. The codec frame size mode (20 ms or 30 ms) is specified by out-of-band means.

8.4.1.2.4 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 [IETF RFC 3951]).

If the 20 ms frame size mode is used, the media endpoint MUST send the "mode" parameter in the SDP "a=fmtp" attribute by copying it 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" stands for preferred 30 ms frame size).

The following is an example of the media representation in SDP for describing iLBC when 20 ms frame size mode is used:

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

Alternatively, if the 30 ms frame size mode is used, the media representation might be:

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

As indicated in the example, when the "mode" parameter in SDP "a=fmtp" attribute is not present, 30 ms frame size mode MUST be applied. Mode negotiation by the media endpoint must be done according to [IETF RFC 3952].

8.4.1.3 BroadVoice16 (BV16)

BV16 was selected as a codec standard suitable for packet-based communication networks. A mathematical description of the codec is available in [ANSI/SCTE 24-21].

BV16 MAY be supported by user equipment and media gateways. BV16 supports a coding rate of 16 kbit/s with a frame size of 5 ms. User equipment and media gateways MUST support 10 ms, 20 ms, and 30 ms packet sizes when BV16 is used.

8.4.1.3.1 Packet loss concealment

For BV16, user equipment and media gateways SHOULD use the method defined in [ANSI/SCTE 24-21] for packet loss concealment.

8.4.1.3.2 Voice activity detection and silence suppression

BV16 does not have an associated VAD mechanism. For BV16, the user equipment and media gateways MAY employ VAD and silence suppression such as discontinuous transmission (DTX) to reduce bandwidth. If silence suppression is used with BV16, the user equipment and media gateways SHOULD transmit silence insertion descriptor frames as specified in Appendix II to [ITU-T G.711].

8.4.1.3.3 Payload header format

User equipment and media gateways MUST support the payload header format specified in [IETF RFC 4298] for BV16. A standard RTP header is used along with one or more frames of BV16 to form the packet. User equipment and media gateways MUST use the codec payload bit packing as specified in [IETF RFC 4298] for BV16. There are no options specific to this payload header format.

8.4.1.3.4 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 [IETF RFC 4298]).

The following is an example of the media representation in SDP for describing BV16 when 20 ms frame size mode is used:

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

8.4.1.4 Adaptive multi rate (AMR)

The AMR codec [b-3GPP TS 26.090] was originally developed for use in GSM cellular systems by ETSI. It has now been chosen for use in 3G cellular systems by 3GPP. It is also a mandatory codec in the 3GPP IP multimedia subsystem (IMS) specifications [ETSI TS 126 235]. IPCablecom needs to provide interworking to cellular systems. Recommending the use of AMR guarantees end-to-end narrow-band codec interoperability between user equipment or media gateways and 3GPP cellular networks. See Appendix II to [ITU-T G.711] for more details.

AMR is a multi-mode codec with eight separate encoding modes at the following bit rates: 4.75, 5.15, 5.9, 6.7, 7.4, 7.95, 10.2 and 12.2 kbit/s. Three of these encoding modes already exist as independent coding standards: the 12.2 kbit/s mode as [b-3GPP TS 46.060], the 7.4 kbit/s mode as [b-TIA/EIA 136-410] and the 6.7 kbit/s mode as [b-ARIB 27H]. All encoding modes use a standard 20 ms frame size.

AMR MAY be supported in user equipment and media gateways. If AMR is supported, all coding rates MUST be supported by user equipment and media gateways.

8.4.1.4.1 Packet loss concealment

User equipment and media gateways supporting AMR SHOULD use the method defined in [ETSI TS 126 091] for packet loss concealment.

8.4.1.4.2 Voice activity detection and silence suppression

User equipment and media gateways that support AMR MUST be capable of supporting voice activity detection (VAD), discontinuous transmission (DTX), silence insertion descriptor (SID), and comfort noise generation (CNG) schemes associated with this codec. This is to allow an IPCablecom UE to handle SID frames and generate CNG in the same fashion as a 3GPP cellular device.

Specifically, user equipment and media gateways implementing AMR MUST:

- support VAD/DTX functions in accordance with [ETSI TS 126 093] and [ETSI TS 126 094]. The choice of VAD1 or VAD2 as outlined in [ETSI TS 126 094] is left to the vendor as this choice does not require any signalling and does not impact on interworking;
- support generation and handling of SID frames in accordance with [b-3GPP 26.911], [ETSI TS 126 093], [ETSI TS 126 101]. (The use of extra SID frame types in [ETSI TS 126 093], i.e., GSM-EFR SID, TDMA-EFR SID, and PDC-EFR SID are NOT required);
- support comfort noise generation in accordance with [ETSI TS 126 092].

8.4.1.4.3 Payload header format

The payload header format for AMR is specified in [IETF RFC 3267]. This RFC outlines a range of supported features and options. A profile of [IETF RFC 3267] outlining the options supported in IMS applications is given in [ETSI TS 126 236]. User equipment and media gateways supporting AMR MUST support the payload header format specified in [IETF RFC 3267] with the options specified in [ETSI TS 126 236]. The implementation requirements for user equipment and media gateways supporting AMR are as follows:

- Bandwidth-efficient versus octet-aligned mode: In octet-aligned mode, all the fields in the RTP payload (payload header, table of contents entries and speech payload) are aligned to octet boundaries. In bandwidth-efficient mode, only the full RTP payload is octet-aligned, so padding bits are only used at the end of the entire RTP payload. It should be noted that certain features such as interleaving, frame CRCs and robust sorting can only be used in conjunction with octet-aligned mode. The use of bandwidth-efficient or octet-aligned mode is signalled by out-of-band means, using the optional 'octet-align' parameter. User equipment and media gateways supporting AMR encode and decode implementations MUST support bandwidth-efficient mode in accordance with [ETSI TS 126 236]. User equipment and media gateways supporting AMR encode and decode implementations MAY support octet-aligned mode.

- Codec mode request (CMR): User equipment and media gateway implementations supporting AMR MUST support the ability to encode and decode ALL codec modes (4.75, 5.15, 5.9, 6.7, 7.4, 7.95, 10.2, 12.2 kbit/s and AMR SID frames), as well as switching to any mode at any 20 ms frame boundary. The codec mode that a near-end AMR decoder prefers to receive is signalled in the CMR field within the payload header sent with AMR frames from the near-end AMR encoder to the far-end AMR decoder. User equipment and media gateway encoders implementing the AMR codec SHOULD follow a received mode request. Using appropriate CMRs, it is quite possible for both media paths in a bidirectional session to be using different codec modes. User equipment and media gateways implementing AMR MUST support the generation and processing of CMR fields as described in [IETF RFC 3267]. The use of CMR itself does not require out-of-band signalling.

  In certain transport networks, the full range of codec modes supported may be restricted to a defined subset. For example, 3GPP usage specified in [b-3GPP TS 44.018] describes an active codec mode set of up to four codec modes to be used on a particular call. The signalling of the active codec mode set is achieved by out-of-band means, using the optional 'mode-set' parameter. In addition, the intervals at which the codec mode may be changed, and whether only neighbouring modes in the active codec mode set can be switched to, are signalled using out-of-band means with the optional "mode-change-period" and "mode-change-neighbor" parameters respectively. User equipment and media gateway AMR encode implementations MAY use "mode-set", "mode-change-period", "mode-change-neighbor". User equipment and media gateway AMR decode implementations MUST support the use of "mode-set", "mode-change-period", "mode-change-neighbor" in accordance with [IETF RFC 3267]. When two or more codec modes are specified with the "mode-set" parameter, "mode-change-period" MUST be set to a value of '2' in order to align with [ETSI TS 126 236].

- Redundant transmission: The RTP payload format specified in [IETF RFC 3267] is capable of sending redundant encodings of speech frames to improve robustness against packet loss. As the primary and redundant version(s) of any speech frame are sent in consecutive packets, this scheme constitutes a subset of the functionality provided by [IETF RFC 2198]. The use of redundant transmission does not require out-of-band signalling. It should be noted that the use of redundancy may substantially increase the end-to-end latency of the speech transmission path. It may also be necessary to adjust QoS flowspecs when redundancy is in use to accommodate the extra media bandwidth required. In accordance with [ETSI TS 126 236], user equipment and media gateway AMR encode implementations MUST NOT use redundant transmission. User equipment and media gateway AMR decode implementations MAY support the processing of payloads with redundant encodings.

- Frame interleaving: Interleaving of AMR encodings can mitigate the effect of packet loss even in bursty channels. [IETF RFC 3267] supports the use of frame interleaving through the transmission of ILL and ILP fields within the payload header indicating the interleaving depth and the interleaving index within any interleaving group respectively. Frame interleaving can only be used when operating in octet-aligned mode. It should be noted that frame interleaving may substantially increase the end-to-end latency of the speech transmission path. Furthermore, interleaving may affect encryption as key changes may need to occur at the boundaries between interleave groups. Frame interleaving is enabled through signalling the 'interleaving' parameter out-of-band. When present, this parameter indicates the maximum number of AMR encodings allowed in an interleaving group. In accordance with [ETSI TS 126 236], frame interleaving MUST NOT be used in user equipment and media gateway AMR implementations.

- Frame CRCs: [IETF RFC 3267] discusses the calculation by the AMR encoder of a CRC on the most sensitive (class A) bits within the AMR speech encoding. The CRC is communicated to the remote decoder by inserting CRC values into the table of contents entries within the RFC 3267 packet. These CRCs are then checked against a recalculation of the CRC by the decoder on the received bits to determine whether any bit errors occurred in transmission. Frame CRCs can only be used when operating in octet-aligned mode. Transmission of frame CRCs is enabled through signalling the "crc" parameter out-of-band. In accordance with [ETSI TS 126 236], frame CRCs MUST NOT be used in user equipment and media gateway AMR implementations.
- Robust sorting: If multiple AMR encodings are packed into one [IETF RFC 3267] payload, the bits within each AMR encoding can be sorted in two ways. With simple sorting, the encodings are packed sequentially one after another. However, when robust sorting is used, the octets within each AMR encoding are interleaved to collect the most sensitive bits towards the start of the payload. This simplifies the use of error detection/correction on the most sensitive bits within each encoding. Robust sorting can only be used when operating in octet-aligned mode. Robust sorting is enabled through signalling the 'robust-sorting' parameter out-of-band. In accordance with [ETSI TS 126 236], simple sorting MUST be supported in user equipment and media gateway AMR implementations and robust sorting MUST NOT be used.

8.4.1.4.4 Session description

Parameters are mapped to SDP in a standard way. When conveying information by SDP, the encoding name MUST be "AMR" (the same as the MIME subtype [IETF RFC 3267]).

The following is an example of the media representation in SDP for describing AMR when bandwidth-efficient mode is used with only four codec modes (4.75, 7.4, 10.2, 12.2 kbit/s) in the active codec mode set and switching between adjacent active modes is permitted only if switches are at least two frame blocks apart:

```
m=audio 49130 RTP/AVP 98
a=rtpmap: 98 AMR/8000
a=mptime: 20
a=fmtp: 98 mode-set=0,4,6,7; mode-change-period=2; mode-change-neighbor=1
```

8.4.1.5 Selectable mode vocoder (SMV)

The SMV codec [ANSI/TIA-127-A] was originally developed by TIA as IS-893 and has now been adopted by 3GPP2 for use in 3G CDMA2000 systems. In addition, it has also been specified for multimedia domain (MMD) applications by 3GPP2. IPCablecom needs to provide interworking to cellular systems. Recommending the use of SMV guarantees end-to-end narrow-band codec interoperability between user equipment or media gateways and 3GPP2 cellular networks.

SMV MAY be supported in user equipment and media gateways. SMV is a variable-bit-rate codec with four possible bit rates: 0.8, 2.0, 4.0 and 8.55 kbit/s. All bit rates use a standard 20 ms frame size. If SMV is supported, all encoding rates MUST be supported.

SMV is a source-controlled codec which is capable of adjusting its encoding rate based on the input signal to the codec using an intelligent rate-determination algorithm (RDA). The particular codec rates used are determined by the operating mode that is chosen. SMV can operate in one of six modes which should not be confused with the encoding mode terminology used for the AMR codec. In the AMR case, codec mode refers to the bit rate used by the codec. In the SMV case, the setting of a mode for the codec determines the bit rates used by the codec in its source-controlled operation and is hence a determinant of overall speech quality. Each mode is capable of choosing any of the codec bit rates. The codec will produce an average bit rate dependent on the input signal and the operating mode. In conversational speech (i.e., approximately 50 percent voice activity factor), the average bit rate for SMV in mode 0 (highest quality) is 3.70 kbit/s while for mode 5 (lowest quality) it is 1.85 kbit/s. The configuration of operating mode is done externally to the codec and is out of scope of this Recommendation. The operating mode of the encoder does not need to be transmitted to the decoder as the SMV decoder does not need additional information other than the codec data frames themselves.

8.4.1.5.1 Packet loss concealment

User equipment and media gateways supporting SMV SHOULD use the frame erasure concealment method defined in [TIA/IS-893] for packet loss concealment.

8.4.1.5.2 Voice activity detection and silence suppression

User equipment and media gateways supporting SMV MUST be capable of supporting voice activity detection (VAD) for rate determination. Support of VAD must be in accordance with [TIA/IS-893]. The decision to use VAD option A or VAD option B is left to the vendor as this choice does not require any signalling and does not impact on interworking. No discontinuous transmission (DTX), silence insertion descriptors (SID), or comfort noise generation (CNG) schemes are currently defined for SMV as SMV is designed for continuous transmission.

8.4.1.5.3 Payload header format

User equipment and media gateways supporting SMV MUST use the payload header format as specified in [IETF RFC 3558]. This RFC outlines a range of supported features and options. As with 3GPP, 3GPP2 has not provided a profile for this RFC.

8.4.1.5.4 Session description

Parameters are mapped to SDP in a standard way. When conveying information by SDP, the encoding name MUST be "SMV" for the interleaved/bundled format and MUST be "SMV0" for the header-free format (the same as the MIME subtypes [IETF RFC 3558]).

The following is an example of the media representation in SDP for describing SMV when the interleaved/bundled format is used with interleaving enabled and a maximum interleaving depth of 3:

```
m=audio 49120 RTP/AVP 97
a=rtpmap: 97 SMV/8000
a=mptime: 20
a=fmtp: 97 maxinterleave=3
```

Alternatively, the following is an example of the media representation in SDP for describing SMV when the header-free format is used:

```
m=audio 49130 RTP/AVP 99
a=rtpmap: 99 SMV0/8000
a=mptime: 20
```

8.4.1.6 Enhanced variable-rate codec (EVRC)

The EVRC codec [ANSI/TIA-127-A] was originally developed by TIA as IS-127 and has now been adopted by 3GPP2 for use in 3G CDMA2000 systems. In addition, it has also been specified for multimedia domain (MMD) applications by 3GPP2. IPCablecom needs to provide interworking to cellular systems. Recommending the use of EVRC guarantees end-to-end narrow-band codec interoperability between user equipment or media gateways and 3GPP2 cellular networks that use EVRC.

EVRC MAY be supported in user equipment and media gateways. EVRC is a variable-bit-rate codec with three possible bit rates: 0.8, 4.0 and 8.55 kbit/s. All bit rates use a standard 20 ms frame size. If EVRC is supported, all encoding rates MUST be supported.

8.4.1.6.1 Packet loss concealment

User equipment and media gateways supporting EVRC SHOULD use the frame erasure concealment method defined in [ANSI/TIA-127-A].

8.4.1.6.2 Voice activity detection and silence suppression

EVRC does not have an associated VAD mechanism. User equipment and media gateways supporting EVRC MAY employ VAD and silence suppression such as discontinuous transmission (DTX) to reduce bandwidth. If silence suppression is used, user equipment and media gateways supporting EVRC SHOULD transmit silence insertion descriptor frames as specified in Appendix II to [ITU-T G.711].

8.4.1.6.3 Payload header format

User equipment and media gateways MUST use the payload header format for EVRC as specified in [IETF RFC 3558]. This RFC outlines a range of supported features and options. As with 3GPP, 3GPP2 has not provided a profile for this RFC.

8.4.1.6.4 Session description

Parameters are mapped to SDP in a standard way. When conveying information by SDP, the encoding name MUST be "EVRC" for the interleaved/bundled format and MUST be "EVRC0" for the header-free format (the same as the MIME subtypes [IETF RFC 3558]).

The following is an example of the media representation in SDP for describing EVRC when the interleaved/bundled format is used with interleaving enabled and a maximum interleaving depth of 3:

```
m=audio 49130 RTP/AVP 98
a=rtpmap: 98 EVRC/8000
a=mptime: 20
a=fmtp: 98 maxinterleave=3
```

Alternatively, the following is an example of the media representation in SDP for describing EVRC when the header-free format is used:

```
m=audio 49150 RTP/AVP 99
a=rtpmap: 99 EVRC0/8000
a=mptime: 20
```

8.4.2 Other feature support and real-time text conversation

The ability to offer a competitive telephone service depends on more than the use of toll quality codecs. The end-to-end narrow-band audio service must compensate for the undesirable effect of echo from the PSTN while adapting for the delay and loss variability inherent in a packet network. The objective is also for an IPCablecom telephony service to transparently support the full plethora of different modem types and other voiceband data devices that have been designed to work with the PSTN with its deterministic TDM behaviour, rather than with packet networks. This means that the IPCablecom network itself cannot be transparent to the audio it carries whenever any legacy service is involved. In addition to PSTN echo cancellation, a number of detectors are required to look for the presence of fax, analogue modem, and hearing-impaired TTY devices, and DTMF digits. Additional information concerning accessibility may be found at: http://www.itu.int/ITU-T/studygroups/com16/accessibility/docs/tacl.pdf.

8.4.2.1 Echo cancellation support

Line echo is created at the telephone interface of a terminal adapter, or at the far-end of the PSTN interface of the media gateway. Specifically, a hybrid transformer (or hybrid) that converts the separate audio transmit and receive signals (four-wire interface) into a single two-wire interface compatible with a standard telephone creates echo back to a remote talker. When the round-trip delay in an audio communication is more than about 20 ms, talker echo can become discernible. An echo canceller is used to remove this echo.

User equipment and media gateways that provide analogue or TDM interfaces MUST provide echo cancellation to remove 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 line echo canceller MUST comply with [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 the case of a terminal adapter at the user equipment, the length of the echo tail is typically short (8 ms or less). For PSTN media gateway applications, however, user equipment and media gateways MUST support echo canceller lengths of 48 ms minimum consistent with [ANSI T1.508]. Vendors MAY choose to differentiate their products by providing longer echo canceller lengths suitable for their applications or other programmable parameters.

If user equipment uses non-standard telephone interface (e.g., four-wire microphone and headset) and the end device has no hybrid, 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.

8.4.2.2 Asymmetrical services support

User equipment and media gateways SHOULD be capable of supporting different codecs for upstream and downstream audio channels. This allows optimization of device resources, network bandwidth, and user service quality.

8.4.2.3 Hearing-impaired services support

Customer premises equipment (CPE) for the hearing-impaired consists of text input/output devices coupled with low-speed voiceband modems. These are commonly referred to as telephone devices for the deaf (TDD) or teletypes (TTY). Typically, these devices interface to the PSTN via an acoustic coupler to a phone or with a regular RJ-11 telephone jack. Any system designed to support TDD/TTY would need to be able to pass DTMF and voiceband modem tones coherently.

User equipment and media gateways used for applications that require support of TDD MUST support detection and transmission of V.18 Annex A hearing-impaired TTY tones. Upon detection of a V.18 Annex A signal, user equipment and media gateways MUST switch the codec to one that supports transmission of V.18 Annex A tones for the remainder of the session. The following codecs are recommended: G.711, G.726 at 32 kbit/s, and G.726 at 40 kbit/s. Depending upon the specific codecs negotiated for the connection, the user equipment MUST reserve and/or commit additional HFC bandwidth to accommodate the requirements of the new codec. Note that no detection or switching is required for cellular text modems, which are inherently compatible with AMR and other cellular codecs.

8.4.2.4 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 the 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., a 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 requirements for DTMF relay for user equipment and media gateways in this clause are applicable only if DTMF transmission is needed by an application.

User equipment and media gateways that need to support DTMF tone transmission MUST support transmission and reception of [IETF RFC 2833] DTMF telephone-events 0-15, which represents the minimum level required for compliance with the RFC. User equipment and media gateways MAY support other telephone-events. Negotiated events MUST be transferred via [IETF RFC 2833] telephone-event packets regardless of the codec specified for the speech.

[IETF RFC 2833] does not specify DTMF tone duration requirements at the termination endpoint, instead relying on DTMF detection accuracy at the origination endpoint. However, [IETF RFC 2833] does reference ITU-T Rec. Q.24 in defining the minimum DTMF tone duration of 40 ms. Additionally, ITU-T Rec. Q.24 includes a duration of less than 40 ms when the DTMF tones may be accepted as DTMF digits (as low as 20 ms). For North American networks, LSSGR [b-Telecordia GR 506] specifies that tone durations more 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.

Considering these industry requirements, user equipment or media gateways with analogue or time division multiplexing (TDM) interfaces MUST detect DTMF tones of 40 ms or more and report their duration relative to the RTP timestamp. User equipment and media gateways MAY detect DTMF digits of duration more than 23 ms, but user equipment or media gateways MUST NOT report DTMF digits when their duration is less than 23 ms. User equipment and media gateways that originate DTMF telephone-events MUST specify a minimum of 50 ms duration in the telephone-event packets. User equipment and media gateways MUST NOT transmit a DTMF telephone-event packet containing a duration field of value zero and SHOULD ignore a received DTMF telephone-event packet containing a duration field of value zero.

The repetition rate of telephone-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. Audio packets being replaced by RFC 2833 packets MUST continue to be suppressed during the redundant transmission of the end-of-event packets. Repetition of the final packet of each event generally ensures satisfactory performance in the event of the occasional lost packet. 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, user equipment and media gateways MUST play out the tone. [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, user equipment and media gateways 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 to send both audio and telephone-events concurrently. To avoid increasing bandwidth requirements, user equipment and media gateways originating DTMF telephone-event packets MUST stop sending audio packets whenever a DTMF digit is detected, with suppression of audio packets continuing until retransmission of the end-of-event packets is complete. 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 user equipment and media gateways according to the duration specified in the event, subject to an optional minimum play-out duration that MAY be provisioned on the user equipment and media gateways. If audio data is also received for the same timestamp period as covered by telephone-event packets, the media endpoint 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 the jitter buffer adapting down to below the event recognition time at the origination point, the telephone-event play out MAY be shortened from the duration specified in the [IETF RFC 2833] telephone-event packet, 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. 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.

8.4.2.4.1 Session description for DTMF relay

The following SDP attributes are applicable to audio service use for DTMF relay:

```
a=<attribute> : <value>

a=rtpmap:<format> <encoding name>/<clock rate>[/<encoding parameters>]
a=rtpmap:<format> telephone-event/8000

a=fmtp:<format> <format specific parameters>
a=fmtp:<format> 0-15
```

Send: One or more of the "a" attribute lines specified below MAY be included.

Receive: One or more of the "a" attribute lines specified below MAY be included and MUST be acted upon accordingly. Attribute values are case-insensitive. Implementations MUST accept the lower case, upper case, and mixed upper/lower case encodings of all attributes.

Note that SDP [IETF RFC 4566] requires unknown attributes to be ignored.

`rtpmap:`

Send: When transmitting an offer, if DTMF relay is supported and desired to be used, then the "rtpmap" attribute with the "telephone-event" encoding name MUST be included. When transmitting an answer, if DTMF relay is supported and desired to be used, and the offer included the "rtpmap" attribute with the "telephone-event" encoding name, then the "rtpmap" attribute with the "telephone-event" encoding name MUST be included and DTMF relay MUST then be used. In all other cases, the "rtpmap" attribute with the "telephone-event" encoding name MUST NOT be included and DTMF relay MUST NOT be used.

Receive: DTMF relay MUST NOT be used if the "rtpmap" attribute with the "telephone-event" encoding name is absent.

`fmtp:`

Send: The "fmtp" attribute MAY be included to indicate which named events a receiver supports. Since all implementations MUST be able to receive events 0 through 15, listing these events is OPTIONAL. If named events other than 0 through 15 are supported and desired, the "fmtp" attribute MUST be included.

Receive: Named events other than 0 through 15 MUST NOT be used if the "fmtp" attribute is absent or if the "fmtp" attribute is present and the named events do not appear in the list.

The following is an example of the media representation in SDP for describing support for DTMF relay with named events 0 through 15:

```
m=audio 49130 RTP/AVP 18 96
a=rtpmap: 96 telephone-event/8000
a=fmtp: 96 0-15
```

8.4.2.5 Fax and modem support

User equipment and media gateways may need to support analogue fax and modem interfaces for several reasons. First, modem equipment is common in residences and customers will continue to use these familiar devices to access their dial-up networks even if they have cable modem access. Second, many SOHO users need to have fax capability. Finally, some low-rate modem standards such as V.22 and V.23 will continue to be used for point-of-sale (POS) and security applications.

For applications that need the support of analogue fax and modem, user equipment and media gateways MUST detect fax/modem signals and, if required by the protocol, signal these detections using the appropriate protocol. The codec at each end is then switched to [ITU-T G.711] for the remainder of the session unless fax relay is to be used as specified in clause 8.4.2.6. 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.168]. After the fax/modem session has completed, echo cancellation MUST be re-enabled.

A more robust solution for supporting modem, TTY, and also available for fax if fax relay is not used, is to employ voiceband data transmission using the method described in [ITU-T V.152] and specified in clause 8.4.2.7. [ITU-T V.152] involves the user equipment and media gateways autonomously switching to a pre-negotiated codec that can accurately relay modem and TTY signals over an IP network. The use of [ITU-T V.152] with [IETF RFC 2198] redundancy also makes the transmission more resilient to packet loss in the network. This is an important feature for [ITU-T V.152] since packet loss causes modems to drop in speed or disconnect. User equipment and media gateways MAY support [ITU-T V.152] with [IETF RFC 2198] redundancy as defined in this Recommendation.

8.4.2.6 Fax relay

IPCablecom needs to provide reliable fax support since fax equipment continues to be used by both residential and business customers. The recommended solution for supporting reliable fax is to employ T.38 fax relay [ITU-T T.38]. T.38 fax relay involves demodulating the T.30 transmission and sending control and image data over the IP network in real-time. At the receiving end, the received data is re-modulated and sent to the fax terminal using another T.30 session in real-time. T.38 fax relay MAY be supported in user equipment and media gateways.

User equipment and media gateways that support [ITU-T T.38] MUST support version 4 of the T.38 specification. This version provides for support of V.34 fax (super group 3) as well as ensuring interoperability with older T.38 implementations for group 3 fax including the original version 1 with which all implementations are required to interoperate. In accordance with version 4, user equipment and media gateways MUST therefore, support the V.27*ter*, V.29 and V.17 modem protocols for page transmission at transfer rates up to 14.4 kbit/s and the V.34 modem protocol for page transmission at transfer rates up to 33.6 kbit/s.

[ITU-T T.38] allows for either TCP or UDP as the transport protocol. Within the UDP transport there is the option to use UDPTL or RTP. User equipment and media gateways MUST support T.38 fax transmission using both UDPTL and RTP. Within [ITU-T T.38], whether using UDPTL or RTP, additional options allow support for redundancy or forward error correction (FEC). User equipment and media gateways MUST support redundancy and MAY support FEC with [ITU-T T.38]. When using redundancy with UDPTL, user equipment and media gateways MUST support a redundancy level of 4 for T.30 control message data and a redundancy level of 1 MUST be used for T.4 phase C data. Redundancy with RTP is based on [IETF RFC 2198] while FEC is based on [IETF RFC 2733]. When using T.38 with RTP, user equipment and media gateways MUST support a redundancy level of 2.

[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. [ITU-T T.38] must be carried over RTP when secure fax relay is required; see clause 6.3.

Annex D to [ITU-T T.38] 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 T.38].

For group 3 rates, user equipment and media gateways MUST be prepared to receive a T.38 UDPTL fax packet of at least 160 bytes. This is based on a 40 ms packetization period and a 14.4 kbit/s data rate. It includes the UDPTL datagram containing T.4 image data with redundancy but without the IP and UDP headers. For V.34 fax at 33.6 kbit/s, user equipment and media gateways MUST be prepared to receive a T.38 UDPTL fax packet of at least 352 bytes for the same 40 ms packetization period.

For QoS and NAT traversal considerations, T.38 fax packets MUST use the same port used by the voice packets for the connection. In addition, user equipment and media gateways MUST send T.38 fax packets at a default 20 ms packetization period in the upstream flow unless another packetization period is negotiated (10/20/30 ms).

Table 8-3 shows the 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 QoS flow characteristics described above.

8.4.2.6.1 Session description for T.38 using UDPTL

The following SDP attributes are applied at the media level and are specific to image service use for [ITU-T T.38] (image/t38):

```
a= <attribute> : <value>
a=T38FaxVersion: <version>
a=T38MaxBitrate: <bitrate>
a=T38FaxRateManagement: <faxratemanagement>
a=T38FaxMaxBuffer: <maxbuffer>
a=T38FaxMaxDatagram: <maxsize>
a=T38FaxUdpEC: <ECmethod>
a=T38FaxFillBitRemoval
a=T38FaxTranscodingMMR
a=T38FaxTranscodingJBIG
```

Send: One or more of the "a" attribute lines specified below MAY be included.

Receive: One or more of the "a" attribute lines specified below MAY be included and MUST be acted upon accordingly. Attribute values are case-insensitive. Implementations MUST accept the lower case, upper case, and mixed upper/lower case encodings of all attributes.

SDP [IETF RFC 4566] requires unknown attributes to be ignored.

NOTE – Some implementations incorrectly use a colon (':') followed by a number (zero or one) after the attributes "T38FaxFillBitRemoval", "T38FaxTranscodingMMR" and "T38FaxTranscodingJBIG". Implementations that receive such erroneous encodings SHOULD interpret the value ":0" as lack of support for the option and all other values as indicating support of the option in question.

`T38FaxVersion:`

As defined in [ITU-T T.38]. The recipient of the offer MUST accept the version or modify the version attribute to be an equal or lower version when transmitting an answer to the initial offer. The recipient of an offer MUST NOT respond with an answer containing a higher version than that which was offered.

Also as defined in [ITU-T T.38]. Early implementations of [ITU-T T.38] equipment may not provide a T.38 version number. In receipt of SDP without the version attribute, the endpoint MUST assume that the version is 0. This is applied in the following discussion on sending and receiving this attribute:

Send: The endpoint MUST indicate the version that it intends to use with the "T38FaxVersion" attribute. However, it MUST NOT indicate a version that is higher than the version received in a RemoteConnectionDescriptor.

Receive: If a RemoteConnectionDescriptor is received and the "T38FaxVersion" attribute is not included, then the endpoint MUST use version 0 of [ITU-T T.38]. If the attribute is included, the endpoint MUST use a version of the specification that is the same or lower than the version indicated.

`T38MaxBitRate:`

Send: The "T38MaxBitRate" attribute SHOULD be included. User equipment and media gateways negotiating T.38 following detection of V.21 flags SHOULD set this parameter to "14400". Following detection of CNG, user equipment and media gateways capable of supporting V.34 fax MAY set this parameter to "33600" in initial offers. The recipient of an offer MUST NOT respond with an answer containing a higher bit rate than that which was offered.

Receive: The "T38MaxBitRate" attribute SHOULD be for bandwidth reservation.

`T38FaxRateManagement:`

Send: The "T38FaxRateManagement" attribute MUST be included and MUST have a value of "transferredTCF" when UDPTL is used. With the value "transferredTCF", TCF is passed end-to-end as opposed to an attribute value of "localTCF" where TCF is generated locally. Note that "localTCF" is only appropriate when a reliable transport such as TCP is used.

Receive: When UDPTL is used, the "T38FaxRateManagement" attribute either MUST be present with a value of "transferredTCF" or it MUST be absent, in which case transferred TCF is assumed. All other values of the attribute MUST be rejected (error code 415 – unsupported media type).

`T38FaxMaxBuffer:`

Send: The "T38FaxMaxBuffer" attribute MUST NOT be included.

Receive: The "T38FaxMaxBuffer" attribute SHOULD be ignored.

`T38FaxMaxDatagram:`

Send: The "T38FaxMaxDatagram" attribute MUST be included. The value indicated MUST NOT be less than 160 bytes. This is based on 40 ms packetization period and a 14'400 bit/s data rate. It includes the UDPTL datagram without the IP and UDP headers.

Receive: Endpoints MUST NOT send a datagram larger than that specified in the "T38FaxMaxDatagram" attribute. Prior to sending any T.38 datagram, the endpoint MUST ensure that it is within the limits defined by this attribute. If the specified "T38FaxMaxDatagram" value is too small to support redundancy for a given datagram, but sufficient to support T.38 without redundancy, then the endpoint MUST send that T.38 datagram without redundancy. If the value is too small to allow the datagram to be sent without redundancy, the endpoint MUST NOT send the T.38 datagram and the command MUST be rejected (error code 415 – unsupported media type).

`T38FaxUdpEC:`

Support for redundancy is mandatory whereas support for forward error correction is optional. Use of either scheme requires negotiation.

Send: The "T38FaxUdpEC" attribute MUST be included. An offer MAY include the value "t38UDPFEC" if FEC is supported. An answer MAY include the value "t38UDPFEC" if FEC is supported and the answer included "t38UDPFEC". Otherwise "t38UDPRedundancy" MUST be sent.

Receive: Redundancy MUST be used if the value of the "T38FaxUdpEC" attribute is "t38UDPRedundancy". If the "T38FaxUdpEC" attribute is "t38UDPFEC" and FEC is supported by the endpoint, then FEC SHOULD be used. If the "T38FaxUdpEC" attribute is "t38UDPFEC" and FEC is not supported, then redundancy MUST be used. If this attribute is not included, the endpoint MUST NOT use redundancy or FEC.

`T38FaxFillBitRemoval:`

Support for fill bit removal is optional and any use of it needs to be negotiated.

Send: When transmitting an offer, if fill bit insertion and removal is supported and desired to be used, then the "T38FaxFillBitRemoval" parameter MUST be included. When transmitting an answer, if fill bit insertion and removal is supported and desired to be used, and the offer included the "T38FaxFillBitRemoval" parameter, then "T38FaxFillBitRemoval" MUST be included and fill bit insertion and removal MUST then be used. In all other cases, the "T38FaxFillBitRemoval" parameter MUST NOT be included and fill bit insertion and removal MUST NOT be used.

Receive: Fill bit insertion and removal MUST NOT be used if the "T38FaxFillBitRemoval" parameter is absent.

`T38FaxTranscodingMMR:`

MMR transcoding does not apply to UDPTL-based T.38.

Send: When UDPTL is being used for T.38, the "T38FaxTranscodingMMR" attribute MUST NOT be included.

Receive: If the "T38FaxTranscodingMMR" attribute is present for UDPTL-based T.38, the command MUST be rejected (error code 415 – unsupported media type).

`T38FaxTranscodingJBIG:`

JBIG transcoding does not apply to UDPTL-based T.38.

Send: When UDPTL is being used for T.38, the "T38FaxTranscodingJBIG" attribute MUST NOT be included.

Receive: If the "T38FaxTranscodingJBIG" attribute is present for UDPTL-based T.38, the command MUST be rejected (error code 415 – unsupported media type).

8.4.2.6.2 Session description for T.38 using RTP

The following SDP attributes are applied at the media level and are applicable to audio service use for T.38 (audio/t38):

```
a=<attribute> : <value>
a=fmtp:<format> <format specific parameters>
a=fmtp:<format> <parameter>=<value>;<parameter>=<value>;…;<parameter>
```

Send: One or more of the "a" attribute lines specified below MAY be included.

Receive: One or more of the "a" attribute lines specified below MAY be included and MUST be acted upon accordingly. Attribute values are case-insensitive. Implementations MUST accept the lower case, upper case, and mixed upper/lower case encodings of all attributes.

Note that SDP [IETF RFC 4566] requires unknown attributes to be ignored.

`fmtp:`

Send: This field MUST be used to provide parameters specific to the "audio/t38" format. At most, one instance of this attribute is allowed, corresponding to the "audio/t38" format per media description. In other words, all of the "audio/t38" format-specific parameters MUST appear on the same "a=fmtp" SDP attribute line. For example:

```
a=fmtp:<format> T38FaxVersion=0;
     T38FaxRateManagement=transferredTCF;
     T38FaxFillBitRemoval; T38FaxTranscodingMMR
```

Receive: When used, the field MUST be used in accordance with [IETF RFC 4566]. Implementations MUST allow for zero or more instances of this attribute corresponding to the "audio/t38" format, each with one or more "audio/t38" format-specific parameters. For example:

```
a=fmtp:<format> T38FaxVersion=0
a=fmtp:<format> T38FaxRateManagement=transferredTCF
a=fmtp:<format> T38FaxFillBitRemoval;T38FaxTranscodingMMR
```

Consider the "audio/t38" format-specific parameters:

`T38FaxVersion:`

As defined in [ITU-T T.38]. The recipient of the offer MUST accept that version or modify the version parameter to be an equal or lower version when transmitting an answer to the initial offer. The recipient of an offer MUST NOT respond with an answer containing a higher version than that which was offered.

Also as defined in [ITU-T T.38]. Early implementations of T.38 equipment may not provide a T.38 version number. In receipt of SDP without the version parameter, the endpoint MUST assume that the version is 0. This is applied in the following discussion on sending and receiving this parameter:

Send: The endpoint MUST indicate the version that it intends to use with the "T38FaxVersion" parameter. However, it MUST NOT indicate a version that is higher than the version received in an offer.

Receive: If an offer is received and the "T38FaxVersion" parameter is not included, then the endpoint MUST use version 0 of [ITU-T T.38]. If the parameter is included, the endpoint MUST use a version of the specification that is the same or lower than the version indicated.

`T38MaxBitRate:`

Send: The "T38MaxBitRate" parameter SHOULD be included. User equipment and media gateways capable of supporting V.34 fax SHOULD set this parameter to "33600" in initial offers. User equipment and media gateways not capable of supporting V.34 fax SHOULD set this parameter to "14400". The recipient of an offer MUST NOT respond with an answer containing a higher bit rate than that which was offered.

Receive: The "T38MaxBitRate" parameter SHOULD be used for bandwidth reservation.

`T38FaxRateManagement:`

Send: The "T38FaxRateManagement" parameter MUST be included and MUST have a value of "transferredTCF" when RTP is used. With the value "transferredTCF", TCF is passed end-to-end as opposed to a parameter value of "localTCF" where TCF is generated locally. Note that "localTCF" is only appropriate when a reliable transport such as TCP is used.

Receive: When RTP is used, the "T38FaxRateManagement" parameter either MUST be present with a value of "transferredTCF" or it MUST be absent, in which case transferred TCF is assumed. All other values of the attribute MUST be rejected (error code 415 – unsupported media type).

`T38FaxMaxBuffer:`

Send: The "T38FaxMaxBuffer" parameter MUST NOT be included.

Receive: The "T38FaxMaxBuffer" parameter SHOULD be ignored.

`T38FaxMaxDatagram:`

Send: The "T38FaxMaxDatagram" attribute MUST be included.

Receive: Endpoints MUST NOT send a datagram larger than that specified in the "T38FaxMaxDatagram" parameter. Prior to sending any T.38 datagram, the endpoint MUST ensure that is within the limits defined by this parameter. If the specified "T38FaxMaxDatagram" value is too small to support redundancy for a given datagram, but sufficient to support T.38 without redundancy, then the endpoint MUST send that T.38 datagram without redundancy. If the value is too small to allow the datagram to be sent without redundancy, the endpoint MUST NOT send the T.38 datagram and the command MUST be rejected (error code 415 – unsupported media type).

`T38FaxFillBitRemoval:`

Support for fill bit removal is optional and any use of it needs to be negotiated.

Send: When transmitting an offer, if fill bit insertion and removal is supported and desired to be used, then the "T38FaxFillBitRemoval" parameter MUST be included. When transmitting an answer, if fill bit insertion and removal is supported and desired to be used, and the offer included the "T38FaxFillBitRemoval" parameter, then "T38FaxFillBitRemoval" MUST be included and fill bit insertion and removal MUST then be used. In all other cases, the "T38FaxFillBitRemoval" parameter MUST NOT be included and fill bit insertion and removal MUST NOT be used.

Receive: Fill bit insertion and removal MUST NOT be used if the "T38FaxFillBitRemoval" parameter is absent.

`T38FaxTranscodingMMR:`

Support for MMR transcoding is optional and any use of it needs to be negotiated.

Send: When transmitting an offer, if MMR transcoding is supported and desired to be used, then the "T38FaxTranscodingMMR" parameter MUST be included. When transmitting an answer, if MMR transcoding is supported and desired to be used, and the offer included the "T38FaxTranscodingMMR" parameter, then the "T38FaxTranscodingMMR" parameter MUST be included and MMR transcoding MUST then be used. In all other cases, the "T38FaxTranscodingMMR" parameter MUST NOT be included and MMR transcoding MUST NOT be used.

Receive: MMR transcoding MUST NOT be used if the "T38FaxTranscodingMMR" parameter is absent.

`T38FaxTranscodingJBIG:`

Support for JBIG transcoding is optional and any use of it needs to be negotiated.

Send: When transmitting an offer, if JBIG transcoding is supported and desired to be used, then the "T38FaxTranscodingJBIG" parameter MUST be included. When transmitting an answer, if JBIG transcoding is supported and desired to be used, and the offer included the "T38FaxTranscodingJBIG" parameter, then the "T38FaxTranscodingJBIG" parameter MUST be included and JBIG transcoding MUST then be used. In all other cases, the "T38FaxTranscodingJBIG" parameter MUST NOT be included and JBIG transcoding MUST NOT be used.

Receive: JBIG transcoding MUST NOT be used if the "T38FaxTranscodingJBIG" parameter is absent.

8.4.2.7 V.152 voiceband data transmission

The recommended method for providing reliable transmission for dial-up modem applications and TTY is to support voiceband data transmission using V.152 procedures [ITU-T V.152] along with [IETF RFC 2198] redundancy. [ITU-T V.152] provides for the pre-negotiation of a codec and payload type expressly for the purpose of carrying voiceband data and defines the triggers that may be used by media gateways to invoke an autonomous switchover to this codec and payload type. The combination of V.152 with redundancy or forward error correction (FEC) allows for modem, fax, and TTY signals to pass through an IP network reliably even when small amounts of packet loss exist. [ITU-T V.152] MAY be supported in user equipment and media gateways.

User equipment and media gateways that support V.152 procedures for voiceband data transmission MUST negotiate use of G.711 as the voiceband data codec and MAY additionally support use of G.726 as a voiceband data codec for TTY and other low-speed modem applications. User equipment and media gateways MUST support a redundancy level of 1 with V.152. User equipment and media gateways MAY support redundancy levels higher than 1, subject to QoS availability and MAY support FEC in addition to redundancy.

Table 8-3 shows the flowspec parameters for 10/20/30 ms voiceband data sessions that can be used in the least-upper-bound calculations for authorization and resource requests (using G.711 as the V.152 codec with a redundancy level of 1).

If V.152 has been negotiated for a connection, user equipment and media gateways MUST transition to voiceband data mode upon detection of any of the following in-band tones:

- CNG (1100 Hz);
- V.21 flags (fax preamble);
- V.18 Annex A tones (TTY);
- V.25 or V.8 answer tone (2100 Hz);
- Bell 103 or 212A answer tone (2225 Hz);
- V.22 Unscrambled binary ones signal (2250 Hz).

User equipment and media gateways MUST transition to V.152 mode on the receipt of packets that are the negotiated payload type for V.152 mode. This ensures that both ends will be switched into V.152 mode as soon as possible.

8.4.2.7.1 Session description for V.152

The following SDP attribute is applicable to audio service use for V.152:

```
a=<attribute> : <value>
a=gpmd:<format> <parameter list>
a=gpmd:<format> "vbd=yes"
a=gpmd:<format> "vbd=no"
```

The following SDP attribute is specific to audio service use for V.152:

```
a=pmft: <modem-fax-transport>
a=pmft: T38
```

Send: One or more of the "a" attribute lines specified below MAY be included.

Receive: One or more of the "a" attribute lines specified below MAY be included and MUST be acted upon accordingly. Attribute values are case-insensitive. Implementations MUST accept the lower case, upper case, and mixed upper/lower case encodings of all attributes.

Note that SDP [IETF RFC 4566] requires unknown attributes to be ignored.

`gpmd:`

The "gpmd" attribute is applied at the media level as defined in [ITU-T V.152]:

Send: The "gpmd" (general-purpose media descriptor) attribute shall be used to associate payload types in a media information ('m') line with VBD mode. The general form of this attribute list is:

```
a=gpmd:<format> <parameter list>
```

In the context of VBD declaration, the <format> MUST be an RTP/AVP payload type. The <parameter list> MUST be a single "parameter=value" pair. This "parameter=value" pair addresses a parameter that is not part of its standard MIME definition. For sessions supporting [ITU-T V.152], the parameter MUST be the Boolean 'vbd' that MUST have the value of 'yes' or 'no'. When set to 'yes' the attribute indicates that the implementation supports VBD mode as described in [ITU-T V.152].

Receive: The field MUST be ignored if it contains a parameter list other than "vbd=yes" or "vbd=no".

Omission of the 'gpmd' attribute with a "vbd=yes" attribute/value pair for any codec in the SDP session description MUST be construed as non support of VBD mode operation as defined in [ITU-T V.152].

`pmft:`

The "pmft" attribute is applied at the session level as defined in [ITU-T V.152]:

Send: When transmitting an offer containing both V.152 and T.38, if T.38 is preferred over V.152, then the "pmft" attribute with "T38" MUST be included. When transmitting an answer containing both V.152 and T.38, if the offer included the "pmft" attribute with "T38", then the "pmft" attribute with "T38" MUST be included. When transmitting an answer containing both V.152 and T.38, if the offer included the "pmft" attribute without "T38", or did not include the "pmft" attribute and the local preference is T.38, then the "pmft" attribute is included, otherwise the "pmft" attribute is not included.

Receive: When receiving an offer containing both V.152 and T.38, if T.38 is supported, and the offer included the "pmft" attribute with "T38", then T.38 MUST be used. When receiving an offer containing both V.152 and T.38, if T.38 is supported, and the offer included the "pmft" attribute without "T38", or did not include the "pmft" attribute and the local preference is T.38, then T.38 is used, otherwise V.152 is used for fax handling. When receiving an answer containing both V.152 and T.38, and the answer included the "pmft" attribute without "T38", or did not include the "pmft" attribute, then V.152 MUST be used. When receiving an answer containing both V.152 and T.38, and the answer included the "pmft" attribute with the "T38" fax transport, then T.38 MUST be used.

In addition to the above attributes associated with V.152, the following SDP attributes are applicable to audio service use for the optional RFC 2833 VBD answer event relay:

```
a=<attribute> : <value>
a=rtpmap:<format> telephone-event/8000
a=fmtp:<format> <format specific parameters>
a=fmtp:<format> 0-15,32-35
```

`fmtp:`

Send: This field MAY be used to indicate the named events for 2100 Hz answer tones (ANS, /ANS, ANSam and /ANSam) that a receiver can handle. If an implementation chooses use of [IETF RFC 2833] for voiceband data answer events it MUST be able to receive all four events, 32 through 35, in addition to the mandatory DTMF events 0 through 15. Support of other voiceband data events is optional.

Receive: An offerer and answerer MAY interpret events 32 through 35 in an answer and offer, respectively, in order to determine whether echo canceller tone disabling is to operate via audio tones or RFC 2833 telephone events.

8.4.2.8 Real-time conversation text

Internationally, PSTN text telephony is standardized in [ITU-T V.18] and [ITU-T T.140]. Protocol for multimedia application text conversation [ITU-T T.140] specifies a common general text conversation format. It introduces real time conversational text as a medium in multimedia communication. The characteristics of the real-time text medium are specified in [ITU-T F.700], and its use in real-time multimedia applications is defined in [ITU-T F.703] and [ITU-T F.724]. The RTP payload format is defined by [IETF RFC 4103].

8.4.3 Codec naming and flow spec parameters for narrow-band codecs

Narrow-band codecs defined in this Recommendation MUST be encoded with string names in the "rtpmap" parameter as shown in Table 8-2.

Table 8-2 – Narrow-band audio codec "rtpmap" parameters

Codec	Literal codec name	rtpmap parameter
G.711 μ-law	PCMU	PCMU/8000
G.711 A-law	PCMA	PCMA/8000
iLBC	iLBC	iLBC/8000
BroadVoice16	BV16	BV16/8000
AMR	AMR	AMR/8000
SMV (interleaved/bundled)	SMV	SMV/8000
SMV (header-free)	SMV0	SMV0/8000
EVRC (interleaved/bundled)	EVRC	EVRC/8000
EVRC (header-free)	EVRC0	EVRC0/8000
RFC 2833 DTMF	telephone-event	Telephone-event/8000
T.38 using RTP	T38	T38/8000
T.140 real-time conversation	t140	t140/1000

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 8-3 specifies 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 8-3 do not include any bandwidth that may be required for media security and the actual values used in resource allocation may need to be adjusted to accommodate IPCablecom security considerations.

For less 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, the bandwidth parameter should contain the least-upper-bound (LUB) of the desired codec bandwidths.

Table 8-3 – Mapping of narrow-band audio codec session description parameters to flowspec

Parameters from session description			Flowspec parameters		Comments
RTP/AVP code	rtpmap	ptime (ms)	Values b, m, M (Note 1) (bytes)	Values r, p (Note 2) (bytes/s)	
0	<none>	10	120	12'000	G.711 μ-law using the payload type defined by IETF
0	<none>	20	200	10'000	
0	<none>	30	280	9'334	
96-127	PCMU/8000	10	120	12'000	G.711 μ-law PCM, 64 kbit/s, default codec
96-127	PCMU/8000	20	200	10'000	
96-127	PCMU/8000	30	280	9'334	
8	<none>	10	120	12'000	G.711 A-law using the payload type defined by IETF
8	<none>	20	200	10'000	
8	<none>	30	280	9'334	
96-127	PCMA/8000	10	120	12'000	G.711 A-law PCM, 64 kbit/s, default codec
96-127	PCMA/8000	20	200	10'000	
96-127	PCMA/8000	30	280	9'334	
96-127	iLBC/8000	20	78	3'900	iLBC, FB-LPC, 15.2 kbit/s, 20 ms frame size with 5 ms look-ahead; 13.3 kbit/s, 30 ms frame with 10 ms look-ahead
96-127	iLBC/8000	30	90	3'000	
96-127	BV16/8000	10	60	6'000	BV16 (narrow-band), 16 kbit/s
96-127	BV16/8000	20	80	4'000	
96-127	BV16/8000	30	100	3'334	
96-127	G726-16/8000	10	60	6'000	
96-127	G726-16/8000	20	80	4'000	
96-127	G726-16/8000	30	100	3'334	
96-127	G726-24/8000	10	70	7'000	
96-127	G726-24/8000	20	100	5'000	
96-127	G726-24/8000	30	130	4'334	
2	<none>	10	80	8'000	G.726-32, identical to G.721, which is assigned payload type 2 by IETF
2	<none>	20	120	6'000	
2	<none>	30	160	5'334	
96-127	G726-32/8000	10	80	8'000	
96-127	G726-32/8000	20	120	6'000	
96-127	G726-32/8000	30	160	5'334	
96-127	G726-40/8000	10	90	9'000	
96-127	G726-40/8000	20	140	7'000	
96-127	G726-40/8000	30	190	6'334	

Table 8-3 – Mapping of narrow-band audio codec session description parameters to flowspec

Parameters from session description			Flowspec parameters		Comments
RTP/AVP code	**rtpmap**	**ptime (ms)**	**Values b, m, M (Note 1) (bytes)**	**Values r, p (Note 2) (bytes/s)**	
15	<none>	10	60	6'000	G.728, assigned payload type 15 by IETF
15	<none>	20	80	4,000	
15	<none>	30	100	3'334	
96-127	G728/8000	10	60	6'000	G.728, LD-CELP, 16 kbit/s
96-127	G728/8000	20	80	4'000	
96-127	G728/8000	30	100	3'334	
18	<none>	10	50	5'000	G.729A, identical to G.729, assigned payload type 18 by IETF
18	<none>	20	60	3'000	
18	<none>	30	70	2'334	
96-127	G729/8000	10	50	5'000	G.729A, CS-ACELP, 8 kbit/s, 10 ms frame size with 5 ms look-ahead
96-127	G729/8000	20	60	3'000	
96-127	G729/8000	30	70	2'334	
96-127	G729E/8000	10	55	5'500	G.729E, CS-ACELP, 11.8 kbit/s, 10 ms frame size with 5 ms look-ahead
96-127	G729E/8000	20	70	3'500	
96-127	G729E/8000	30	85	2'834	
96-127	AMR/8000	20	54	2'700	AMR at 4.75 kbit/s
96-127	AMR/8000	20	55	2'750	AMR at 5.15 kbit/s
96-127	AMR/8000	20	57	2'850	AMR at 5.9 kbit/s
96-127	AMR/8000	20	59	2'950	AMR at 6.7 kbit/s
96-127	AMR/8000	20	60	3'000	AMR at 7.4 kbit/s
96-127	AMR/8000	20	62	3'100	AMR at 7.95 kbit/s
96-127	AMR/8000	20	67	3'350	AMR at 10.2 kbit/s
96-127	AMR/8000	20	72	3'600	AMR at 12.2 kbit/s
96-127	SMV0/8000 (Note 3)	20	42	2'100	SMV at 0.8 kbit/s
96-127	SMV0/8000 (Note 3)	20	45	2'250	SMV at 2.0 kbit/s
96-127	SMV0/8000 (Note 3)	20	50	2'500	SMV at 4.0 kbit/s
96-127	SMV0/8000 (Note 3)	20	62	3'100	SMV at 8.55 kbit/s
96-127	EVRC0/8000 (Note 3)	20	42	2'100	EVRC at 0.8 kbit/s
96-127	EVRC0/8000 (Note 3)	20	50	2'500	EVRC at 4.0 kbit/s
96-127	EVRC0/8000 (Note 3)	20	62	3'100	EVRC at 8.55 kbit/s
96-127	red/8000	10	205	20'500	RFC 2198 redundancy used for G.711, used as a V.152 codec with redundancy of level 1
96-127	red/8000	20	365	18'250	
96-127	red/8000	30	525	17'500	

Table 8-3 – Mapping of narrow-band audio codec session description parameters to flowspec

Parameters from session description			Flowspec parameters		Comments
RTP/AVP code	rtpmap	ptime (ms)	Values b, m, M (Note 1) (bytes)	Values r, p (Note 2) (bytes/s)	
96-127	T38/8000	10	59	5'900	T.38 RTP group 3 fax relay, no redundancy
96-127	T38/8000	20	77	3'850	
96-127	T38/8000	30	95	3'167	
96-127	T38/8000	10	83	8'300	T.38 RTP V.34 fax relay, no redundancy
96-127	T38/8000	20	125	6'250	
96-127	T38/8000	30	167	5'567	
96-127	T38/8000	10	88	8'800	T.38 RTP group 3 fax relay redundancy level 1
96-127	T38/8000	20	124	6'200	
96-127	T38/8000	30	160	5'334	
96-127	T38/8000	10	136	13'600	T.38 RTP V.34 fax relay redundancy level 1
96-127	T38/8000	20	220	11'000	
96-127	T38/8000	30	304	10'134	
96-127	T38/8000	10	112	11'200	T.38 RTP group 3 fax relay redundancy level 2
96-127	T38/8000	20	166	8'300	
96-127	T38/8000	30	220	7'334	
96-127	T38/8000	10	184	18'400	T.38 RTP V.34 fax relay redundancy level 2
96-127	T38/8000	20	310	15'500	
96-127	T38/8000	30	436	14'534	
N/A	N/A	10	62	6'200	T.38 UDPTL group 3 fax relay (without redundancy)
N/A	N/A	20	80	4'000	
N/A	N/A	30	98	3'267	
N/A	N/A	10	86	8'600	T.38 UDPTL V.34 fax relay (without redundancy)
N/A	N/A	20	128	6'400	
N/A	N/A	30	170	5'667	
N/A	N/A	10	80	8'000	T.38 UDPTL group 3 fax relay (with T.4 redundancy level 1)
N/A	N/A	20	116	5'800	
N/A	N/A	30	152	5'067	
N/A	N/A	10	128	12'800	T.38 UDPTL V.34 fax relay (with T.4 redundancy level 1)
N/A	N/A	20	212	10'600	
N/A	N/A	30	296	9'867	
NOTE 1 – "b" is bucket depth (bytes). "m" is minimum policed unit (bytes). "M" is maximum datagram size (bytes)					
NOTE 2 – "r" is bucket rate (bytes/s). "p" is peak rate (bytes/s).					
NOTE 3 – Header-free payload header format is assumed for the SMV and EVRC codecs.					

8.5 Wideband codec specification

For the purpose of this Recommendation, wideband codecs are defined as those which operate on audio signals band-pass filtered to a frequency range of 50 Hz-7 kHz [ITU-T G.712] and sampled at 16'000 samples/s. Similar to narrow-band, the input to the codec will generally be in the form of 16-bit uniformly quantized samples with at least 14 bits of dynamic range. A comparison of known wideband codecs is provided in Appendix II.

8.5.1 Supported wideband codecs

The following clauses describe every wideband codec supported in IPCablecom. Whether a wideband codec is mandatory, recommended or optional depends on the application for which it is used. Therefore, the normative status of each codec is indicated in the associated application capability documents. However, if a particular codec is supported for an application, all the requirements for that codec as specified in this clause MUST be met.

8.5.1.1 G.722

[ITU-T G.722] is the earliest international standard on wideband speech coding. Today, it is mainly used in video teleconferencing systems.

G.722 MAY be supported in user equipment and media gateways. G.722 is a multi-rate wideband speech codec for 16 kHz sampled signals. It has three selectable bit rates: 48, 56 and 64 kbit/s. The 48 kbit/s version of G.722 produces medium-quality wideband speech, and the 56 and 64 kbit/s versions produce good- to high-quality wideband speech. User equipment and media gateways using the G.722 codec MUST support 64 kbit/s and SHOULD support 56 and 48 kbit/s.

8.5.1.1.1 Packet loss concealment

G.722 does not have an associated PLC mechanism. For the G.722 codec, user equipment and media gateways SHOULD employ a PLC mechanism of the vendor's choice.

8.5.1.1.2 Voice activity detection and silence suppression

G.722 does not have an associated VAD mechanism. For use with the G.722 codec, user equipment and media gateways SHOULD employ VAD and silence suppression (DTX) to reduce bandwidth using a mechanism of the vendor's choice. If silence suppression is used with G.722, user equipment and media gateways SHOULD transmit silence insertion descriptor frames as specified in Appendix II to [ITU-T G.711].

8.5.1.1.3 Payload header format

No specific payload header format is specified. Standard RTP usage applies as per [IETF RFC 3550] and [IETF RFC 3551].

8.5.1.1.4 Session description

Parameters are mapped to SDP in a standard way. When conveying information by SDP, the encoding name MUST be "G722". G.722 has a static payload type of 9 as specified in [IETF RFC 3551].

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

```
m=audio 3456 RTP/AVP 9
a=mptime: 20
```

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

```
m=audio 3456 RTP/AVP 99
a=rtpmap: 99 G722-64/16000
a=mptime: 20
```

8.5.1.2 BroadVoice32 (BV32)

BroadVoice32 (BV32) MAY be supported in user equipment and media gateways. BV32 is a wideband speech codec for 16 kHz sampled signals. BV32 is a 32 kbit/s, wideband speech codec. BV32 is very similar to BV16 in terms of the coding algorithm.

8.5.1.2.1 Packet loss concealment

BV32 has an associated PLC mechanism similar to BV16. User equipment and media gateways SHOULD use the method defined for BV32.

8.5.1.2.2 Voice activity detection and silence suppression

BV32 does not have an associated VAD mechanism. For the BV32 codec, user equipment and media gateways MAY employ VAD and silence suppression (DTX) to reduce bandwidth. If silence suppression is used with the BV32 codec then user equipment and media gateways SHOULD transmit silence insertion descriptor frames as specified in Appendix II to [ITU-T G.711].

8.5.1.2.3 Payload header format

User equipment and media gateways MUST support the payload header format for BV32 as specified in [IETF RFC 4298]. A standard RTP header is used along with one or more frames of BV32 to form the packet. Any user equipment and media gateway implementation of BV32 MUST use the codec payload bit packing as specified in [IETF RFC 4298]. There are no options specific to this payload header format.

8.5.1.2.4 Session description

Parameters are mapped to SDP in a standard way. When conveying information by SDP, the encoding name MUST be "BV32" [IETF RFC 4298].

The following is an example of the media representation in SDP for describing BV32 when 20 ms frame size mode is used:

```
m=audio 3456 RTP/AVP 98
a=rtpmap: 98 BV32/16000
a=mptime: 20
```

8.5.1.3 Adaptive multi-rate – Wideband (AMR-WB/[ITU-T G.722.2])

The AMR-WB codec [ETSI TS 126 190] was originally developed for use in GSM cellular systems by ETSI. It has also been standardized by ITU-T as [ITU-T G.722.2]. AMR-WB has been chosen for use in 3G cellular systems by 3GPP. It is also a mandatory coder in the 3GPP IP multimedia subsystem (IMS) specifications [ETSI TS 126 235]. IPCablecom needs to provide interworking to cellular systems. Recommending the use of AMR-WB guarantees end-to-end wideband codec interoperability between user equipment or media gateways and 3GPP cellular networks. See [b-ETSI TR 102 493] for more details.

AMR-WB is a variable-bit-rate wideband speech codec for 16 kHz sampled signals. It has nine selectable encoding modes at the following bit rates: 6.60, 8.85, 12.65, 14.25, 15.85, 18.25, 19.85, 23.05 and 23.85 kbit/s. Except for the lowest two modes of 6.60 and 8.85 kbit/s, AMR-WB gives good- to high-quality speech in other modes. When used in 3G GSM networks, the bit rate of

AMR-WB is controlled by the condition of the transmission channel. All encoding modes use a standard 20 ms frame size.

AMR-WB MAY be supported in user equipment and media gateways. User equipment and media gateways supporting AMR-WB MUST support all coding modes.

8.5.1.3.1 Packet loss concealment

User equipment and media gateways supporting AMR-WB codec SHOULD use the method defined in [ETSI TS 126 191] for packet loss concealment.

8.5.1.3.2 Voice activity detection and silence suppression

User equipment and media gateways that support AMR-WB MUST be capable of supporting voice activity detection (VAD), discontinuous transmission (DTX), silence insertion descriptor (SID) and comfort noise generation (CNG) schemes associated with this codec. This is to allow user equipment and media gateways to handle SID frames and generate CNG in the same fashion as a 3GPP cellular device.

Specifically, user equipment and media gateways implementing AMR-WB MUST:

- Support VAD/DTX functions in accordance with [ETSI TS 126 193] and [ETSI TS 126 194].
- Support generation and handling of SID frames in accordance with [ETSI TS 126 191], [ETSI TS 126 193] and [ETSI TS 126 201].
- Support comfort noise generation in accordance with [ETSI TS 126 192].

8.5.1.3.3 Payload header format

The payload header format is specified in [IETF RFC 3267]. This RFC outlines a range of supported features and options. User equipment and media gateways MUST adhere to the formats specified in [IETF RFC 3267]. A profile of RFC 3267 outlining the options supported in IMS applications is given in the 3GPP specification [ETSI TS 126 236] which contains further recommendations for conversational usage of AMR-WB over a packet-switched network, e.g., VoIP over cable. The implementation requirements for user equipment and media gateways supporting AMR-WB are as follows:

- Bandwidth-efficient versus octet-aligned mode: In octet-aligned mode, all the fields in the RTP payload (payload header, table of contents entries and speech payload) are aligned to octet boundaries. In bandwidth-efficient mode, only the full RTP payload is octet-aligned, so padding bits are only used at the end of the entire RTP payload. It should be noted that certain features such as interleaving, frame CRCs and robust sorting can only be used in conjunction with octet-aligned mode. The use of bandwidth-efficient or octet-aligned mode is signalled by out-of-band means, using the optional "octet-align" parameter. User equipment and media gateways supporting AMR-WB encode and decode implementations MUST support bandwidth-efficient mode in accordance with [ETSI TS 126 236]. User equipment and media gateways supporting AMR-WB encode and decode implementations MAY support octet-aligned mode.
- Codec mode request (CMR): User equipment and media gateways supporting AMR-WB MUST support the ability to encode and decode ALL codec modes (6.60, 8.85, 12.65, 14.25, 15.85, 18.25, 19.85, 23.05 and 23.85 kbit/s and AMR-WB SID frames) as well as switching to any mode at any 20 ms frame boundary. The codec mode that a near-end AMR-WB decoder prefers to receive is signalled in the CMR field within the payload header sent with AMR-WB frames from the near-end AMR-WB encoder to the far-end AMR-WB decoder. An encoder SHOULD follow a received mode. Using appropriate CMRs, it is quite possible for both media paths in a bidirectional session to be using different codec modes. User equipment and media gateways supporting AMR-WB MUST

support the generation and processing of CMR fields as described in [IETF RFC 3267]. The use of CMR itself does not require out-of-band signalling.

In certain transport networks, the full range of codec modes supported may be restricted to a defined subset. For example, 3GPP usage specified in [b-3GPP TS 44.018] describes an active codec mode set of up to four codec modes to be used on a particular call. The signalling of the active codec mode set is achieved by out-of-band means, using the optional "mode-set" parameter. In addition, the intervals at which the codec mode may be changed, and whether only neighbouring modes in the active codec mode set can be switched to, are signalled using out-of-band means, with the optional "mode-change-period" and "mode-change-neighbor" parameters respectively. User equipment and media gateways supporting AMR-WB encode implementations MAY use "mode-set", "mode-change-period", "mode-change-neighbor". User equipment and media gateways supporting AMR-WB decode implementations MUST support the use of "mode-set", "mode-change-period", "mode-change-neighbor" in accordance with [IETF RFC 3267]. When two or more codec modes are specified with the "mode-set" "parameter", "mode-change-period" MUST be set to a value of 2 in order to align with [ETSI TS 126 236].

- Redundant transmission: The RTP payload format specified in [IETF RFC 3267] is capable of sending redundant encodings of speech frames to improve robustness against packet loss. As the primary and redundant version(s) of any speech frame are sent in consecutive packets, this scheme constitutes a subset of the functionality provided by [IETF RFC 2198]. The use of redundant transmission does not require out-of-band signalling. It should be noted that the use of redundancy may substantially increase the end-to-end latency of the speech transmission path. It may also be necessary to adjust flowspecs when redundancy is in use to accommodate the extra media bandwidth required. In accordance with [ETSI TS 126 236], AMR-WB encode implementations MUST NOT use redundant transmission. AMR-WB decode implementations MAY support the processing of payloads with redundant encodings.
- Frame interleaving: Interleaving of AMR-WB encodings can mitigate the effect of packet loss even in bursty channels. [IETF RFC 3267] supports the use of frame interleaving through the transmission of ILL and ILP fields within the payload header indicating, respectively, the interleaving depth and the interleaving index within any interleaving group. Frame interleaving can only be used when operating in octet-aligned mode. Note that frame interleaving may substantially increase the end-to-end latency of the speech transmission path. Furthermore, interleaving may affect encryption as key changes may need to occur at the boundaries between interleave groups. Frame interleaving is enabled through signalling the "interleaving" parameter out-of-band. When present, this parameter indicates the maximum number of AMR-WB encodings allowed in an interleaving group. In accordance with [ETSI TS 126 236], frame interleaving MUST NOT be used in AMR-WB implementations.
- Frame CRCs: [IETF RFC 3267] discusses the calculation by the AMR-WB encoder of a CRC on the most sensitive (class A) bits within the AMR-WB speech encoding. The CRC is communicated to the remote decoder by inserting CRC values into the table of contents entries within the [IETF RFC 3267] packet. These CRCs are then checked against a recalculation of the CRC by the decoder to determine whether any bit errors occurred in transmission. Frame CRCs can only be used when operating in octet-aligned mode. Transmission of frame CRCs is enabled through signalling the "crc" parameter out-of-band. In accordance with [ETSI TS 126 236], frame CRCs MUST NOT be used in AMR-WB implementations.

- Robust sorting: If multiple AMR-WB encodings are packed into one [IETF RFC 3267] payload, the bits within each AMR-WB encoding can be sorted in two ways. With simple sorting, the encodings are packed sequentially one after another. However, when robust sorting is used, the octets within each AMR-WB encoding are interleaved to collect the most sensitive bits towards the start of the payload. This simplifies the use of error detection/correction on the most sensitive bits within each encoding. Robust sorting can only be used when operating in octet-aligned mode. Robust sorting is enabled through signalling the "robust-sorting" parameter out-of-band. In accordance with [ETSI TS 126 236], simple sorting MUST be supported in AMR-WB implementations and robust sorting MUST NOT be used.

8.5.1.3.4 Session description

Parameters are mapped to SDP in a standard way. When conveying information by SDP, the encoding name MUST be "AMR-WB" (the same as the MIME subtype [IETF RFC 3267]).

The following is an example of the media representation in SDP for describing AMR-WB:

```
m=audio 49120 RTP/AVP 97
a=rtpmap: 97 AMR-WB/16000
a=fmtp:97 mode-change-period=2; mode-change-neighbor=1
a=maxptime: 20
```

According to [IETF RFC 3267] this example specifies that codec mode changes shall be performed in integer multiples of 40 ms, only changes to neighbouring modes are allowed, and that each packet shall represent 20 ms of speech (one codec frame of AMR-WB). This example is consistent with [IETF RFC 3267].

8.5.1.4 Variable-rate multi-mode – Wideband (VMR-WB)

VMR-WB [TIA-1016-A] is the wideband speech codec standardized by 3GPP2 for use in 3G CDMA2000 systems. In addition, it has been specified for multimedia domain (MMD) applications by 3GPP2. IPCablecom needs to provide interworking to cellular systems. Recommending the use of VMR-WB guarantees end-to-end wideband codec interoperability between user equipment or media gateways and 3GPP2 cellular networks.

VMR-WB MAY be supported in user equipment and media gateways. VMR-WB is a variable-rate codec that is source-controlled, i.e., it is capable of adjusting its encoding rate based on the input signal to the codec. The particular codec rates used are determined by the operating mode that is chosen. VMR-WB can operate in one of five modes that should not be confused with the encoding mode terminology used for the AMR-WB codec. In the AMR-WB case, codec mode refers to the bit rate used by the codec. In the VMR-WB case, the setting of a mode for the codec determines the bit rates used by the codec in its source-controlled operation and is hence a determinant of overall speech quality.

Each codec mode of VMR-WB is capable of choosing between several codec bit rates. Modes 0, 1, 2 and 3 communicate at 13.3, 6.2, 2.7 and 1.0 kbit/s. Mode 4 communicates at 8.55, 4.0 and 0.8 kbit/s. The VMR-WB modes are determined by the service option used within the CDMA2000 network. For service option 62, modes 0, 1 and 2 are supported. Service option 63 supports mode 4. In addition, mode 3 allows interoperable operation with AMR-WB. However, only service options 62 and 63 are specified for CDMA2000 terminals that support VMR-WB. All bit rates use a standard 20 ms frame size. User equipment and media gateways supporting VMR-WB MUST support all encoding rates within service options 62 and 63.

The codec produces an average bit rate dependent on the input signal and the operating mode. In conversational speech (i.e., approximately 50% voice activity factor), the average bit rate for VMR-WB in mode 0 (highest quality) is 9.1 kbit/s while for modes 1 and 2 it is 7.6 kbit/s and 6.2 kbit/s respectively. The configuration of the operating mode is done externally to the codec and is out of scope of this Recommendation. The operating mode of the encoder does not need to be transmitted to the decoder as the VMR-WB decoder does not need additional information other than the codec data frames themselves.

8.5.1.4.1 Packet loss concealment

User equipment and media gateways implementing VMR-WB SHOULD use the method defined in VMR-WB specification for packet loss concealment.

8.5.1.4.2 Voice activity detection and silence suppression

User equipment and media gateways that support VMR-WB MUST be capable of supporting voice activity detection (VAD), discontinuous transmission (DTX), silence insertion descriptor (SID), and comfort noise generation (CNG) schemes associated with these codecs. This is to allow an IPCablecom UE to handle SID frames and generate CNG in the same fashion as a 3GPP2 cellular device.

8.5.1.4.3 Payload header formats

User equipment and media gateways that support VMR-WB MUST adhere to the formats specified in [IETF RFC 4348] and [IETF RFC 4424]. However, 3GPP2 has not yet defined a specific profile.

8.5.1.4.4 Session description

Parameters are mapped to SDP in a standard way. When conveying information by SDP, the encoding name MUST be "VMR-WB" (the same as the MIME subtype [TIA-1016-A]).

The following is an example of the media representation in SDP for describing VMR-WB:

```
m=audio 49120 RTP/AVP 98
a=rtpmap: 98 VMR-WB/16000
a=maxptime: 20
```

According to [TIA-1016-A], this example specifies that each packet shall represent 20 ms of speech (one codec frame of VMR-WB). By default, all operating modes in the set of 0 to 3 are allowed, the number of channels shall be one, header-free payload format shall be used (maximum bandwidth efficiency), interleaving shall not be used, and the DTX of VMR-WB shall not be used.

Note that VMR-WB and AMR-WB can enter tandem-free operation in a limited mode. Basically, both must use octet-aligned mode of operation (not bandwidth efficient) and only codec modes 0, 1 and 2 of the AMR-WB codec must be used (6.6 kbit/s, 8.85 kbit/s and 12.65 kbit/s, respectively). An example of such an offer-answer exchange between a CDMA2000 and a WCDMA terminal can be found in [TIA-1016-A].

8.5.2 Feature support

Unlike narrow-band codecs, wideband codecs are not used on connections to the PSTN. They are, therefore, not required to include special features and audio detectors to support legacy PSTN.

8.5.2.1 Fax and modem support

Fax and modem operation is not applicable to wideband codecs, and fax and modem detectors are not inserted on wideband codec connections. Wideband-capable user equipment and media gateways providing analogue POTS interfaces or PSTN interfaces are not expected to allow use of wideband codecs on connections that use these interfaces.

8.5.2.2 Echo cancellation support

Wideband audio terminals are inherently four-wire with separate transmit and receive signal directions. As such, the traditional two-wire to four-wire hybrid that exists on the POTS interface in the PSTN does not exist in wideband audio paths. Without the signal echo that results from a hybrid, line echo cancellation is not required with wideband codecs.

Where a wideband audio terminal establishes a call to the PSTN, the media gateway connecting to the PSTN is required to provide echo cancellation. However, such a call would be restricted to narrow-band codecs for which echo cancellation requirements are specified in clause 8.4.2.1.

In a media endpoint where a non-standard telephone interface is used, e.g., a four-wire microphone and headset connected to a PC, or a loudspeaker telephone with built-in microphone, acoustic echo can be present. In this case, acoustic echo cancellation may be necessary, and vendors implementing these products are expected to employ acoustic echo cancellation.

8.5.2.3 Asymmetrical services support

The requirement specified in clause 8.4.2.2 for user equipment and media gateways to support different codecs for upstream and downstream audio channels also applies to wideband codecs.

8.5.2.4 Hearing-impaired and real-time conversation services support

Acoustically-coupled text telephone devices may be used with wideband codecs. Therefore, wideband user equipment and media gateways MUST support detection of hearing-impaired tones, as defined in Annex A to [ITU-T V.18], in the same way as described in clause 8.4.2.3.

Upon detection of a V.18 Annex A signal, wideband codecs that cannot faithfully transfer the V.18 Annex A tones MUST be switched to a codec that supports transmission of these tones for the remainder of the session. These codecs are recommended: G.711, G.726 at 32 kbit/s; and G.726 at 40 kbit/s. Depending upon the specific codecs negotiated for the connection, user equipment and media gateways MUST reserve and/or commit additional HFC bandwidth to accommodate the requirements of the new codec. Note that no detection or switching is required for cellular text modems which are inherently compatible with the AMR-WB and other cellular codecs.

If real-time text conversation service is required by the IPCablecom applications, wideband user equipment and media gateways MUST also support real-time text as defined in [ITU-T T.140] in the same way as described in clause 8.4.2.8.

8.5.2.5 DTMF relay

Though a legacy PSTN subscriber signalling system, DTMF may continue to find application in wideband telephony because of its widespread use in applications such as feature programming, IVR, voice mail, and telephone conference control. As some of these applications evolve to wideband audio, so the need to provide reliable DTMF transmission will carry forward from narrow-band to wideband. DTMF relay is specified in clause 8.4.2.4 for this purpose and is applicable to wideband codecs as well as narrow-band codecs.

User equipment and media gateways supporting wideband codecs MUST relay DTMF digits using RFC 2833 telephone-event packets per the requirements specified in clause 8.4.2.4. The timestamp unit used for telephone-event packets MUST match the timestamp unit used in the underlying audio. When a wideband audio codec is in use the timestamp unit is 62.5 μs, corresponding to 16'000 samples/s. In this case the SDP used to negotiate DTMF relay for a wideband session MUST specify telephone-event/16000 as a media attribute.

8.5.2.6 Fax relay and V.152

Fax or modem operation is not applicable to wideband codecs and fax detectors are not inserted on wideband codec connections. Wideband-capable user equipment and media gateways providing analogue POTS interfaces or PSTN interfaces are not expected to allow use of wideband codecs on connections that use these interfaces. Multi-function user equipment or media gateways that integrate both fax, modem and wideband codecs, should initiate fax calls as T.38, V.152 or G.711.

8.5.3 Codec naming and flow spec parameters for wideband codecs

The wideband codecs defined in this Recommendation MUST be encoded with the following string names as defined in Table 8-4 in the "rtpmap" parameter:

Table 8-4 – Wideband audio codec "rtpmap" parameters

Codec	Literal codec name	"rtpmap" parameter
BroadVoice32	BV32	BV32/16000
G.722 at 48 kbit/s	G722-48	G722-48/8000
G.722 at 56 kbit/s	G722-56	G722-56/8000
G.722 at 64 kbit/s	G722-64	G722-64/8000
AMR-WB	AMR-WB	AMR-WB/16000
VMR-WB	VMR-WB	VMR-WB/16000

Unknown "rtpmap" parameters SHOULD be ignored if they are received by user equipment and media gateways.

For every defined codec, whether it is represented in SDP as a static or dynamic payload type, Table 8-5 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 8-5 do not include any bandwidth that may be required for media security and the actual values used in resource allocation may need to be adjusted to accommodate IPCablecom security considerations.

For less 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, the bandwidth parameter should contain the least-upper-bound (LUB) of the desired codec bandwidths.

Table 8-5 – Mapping of wideband audio codec session description parameters to flowspec

Parameters from session description			Flowspec parameters		Comments
RTP/AVP code	rtpmap	ptime (ms)	Values b, m, M (Note 1) (bytes)	Values r, p (Note 2) bytes/s	
96-127	BV32/16000	10	80	8'000	BV32 (wideband) 32 kbit/s
96-127	BV32/16000	20	120	6'000	
96-127	BV32/16000	30	160	5'334	
9	<none>	10	120	12'000	G.722 at 64 kbit/s using the Payload Type defined by IETF
9	<none>	20	200	10'000	
9	<none>	30	280	9'334	
96-127	G722-48/16000	10	100	10'000	G.722 at 48 kbit/s using dynamic payload type
96-127	G722-48/16000	20	160	8'000	
96-127	G722-48/16000	30	220	7'333	
96-127	G722-56/16000	10	110	11'000	G.722 at 56 kbit/s using dynamic payload type
96-127	G722-56/16000	20	180	9'000	
96-127	G722-56/16000	30	250	8'333	
96-127	G722-64/16000	10	120	12'000	G.722 at 64 kbit/s using dynamic payload type
96-127	G722-64/16000	20	200	10'000	
96-127	G722-64/16000	30	280	9'333	
96-127	AMR-WB/16000	20	58	2'900	AMR-WB at 6.6 kbit/s
96-127	AMR-WB/16000	20	64	3'200	AMR-WB at 8.85 kbit/s
96-127	AMR-WB/16000	20	73	3'650	AMR-WB at 12.65 kbit/s
96-127	AMR-WB/16000	20	77	3'850	AMR-WB at 14.25 kbit/s
96-127	AMR-WB/16000	20	81	4'050	AMR-WB at 15.85 kbit/s
96-127	AMR-WB/16000	20	87	4'350	AMR-WB at 18.25 kbit/s
96-127	AMR-WB/16000	20	91	4'550	AMR-WB at 19.85 kbit/s
96-127	AMR-WB/16000	20	99	4'950	AMR-WB at 23.05 kbit/s
96-127	AMR-WB/16000	20	101	5'050	AMR-WB at 23.85 kbit/s
96-127	VMR-WB/16000 (Note 3)	20	42	2'100	VMR-WB at 0.8 kbit/s
96-127	VMR-WB/16000 (Note 3)	20	43	2'150	VMR-WB at 1.0 kbit/s
96-127	VMR-WB/16000 (Note 3)	20	47	2'350	VMR-WB at 2.7 kbit/s
96-127	VMR-WB/16000 (Note 3)	20	50	2'500	VMR-WB at 4.0 kbit/s
96-127	VMR-WB/16000 (Note 3)	20	56	2'800	VMR-WB at 6.2 kbit/s
96-127	VMR-WB/16000 (Note 3)	20	62	3'100	VMR-WB at 8.55 kbit/s
96-127	VMR-WB/16000 (Note 3)	20	74	3'700	VMR-WB at 13.3 kbit/s

NOTE 1 – 'b' is bucket depth (bytes). 'm' is minimum policed unit (bytes). 'M' is maximum datagram size (bytes).

NOTE 2 – 'r' is bucket rate (bytes/s). 'p' is peak rate (bytes/s).

NOTE 3 – The header-free payload header format is assumed for the VMR-WB codec.

8.6 Super-wideband codec specifications

Super-wideband codecs are commonly defined as those which operate on audio signals of a minimum nominal frequency range of 20 Hz – 14 kHz sampled at 32'000 samples per second or higher. In this Recommendation, however, the codecs listed in this clause may support lower sampling rate options that translate into frequency responses below this range, but not below a range of 20 Hz – 7 kHz. Similar to narrow-band and wideband, the input to the codec will generally be in the form of 16-bit uniformly quantized samples, but for super-wideband the input is expected to have at least 16 bits of dynamic range.

8.6.1 Supported super-wideband codecs

The following subclauses describe super-wideband codecs specified for use within IPCablecom2. Whether a super-wideband codec is mandatory, recommended or optional depends on the application for which it is used. Therefore, the normative status of each codec is indicated in the associated application capability documents. However, if a particular codec is supported for an application, all the requirements for that codec as specified in this clause MUST be met.

8.6.1.1 AAC

The advanced audio coding (AAC) is specified by MPEG-2 Part 7 [ISO/IEC 13818-7] as the next-generation audio codec to MP3, which is specified by MPEG-1 and MPEG-2 Part 3. AAC is specified with additional tools and options as the general audio codec within MPEG-4 Part 3 [ISO/IEC 14496-3]. Because of its use in the Apple iPod™[1] and other "MP3" players, and by Apple's iTunes, where it has replaced the older MP3 codec, AAC has become today's most widely used full-bandwidth audio codec over the Internet. For this reason, it is recommended for support within IPCablecom2.

AAC is a family of audio codecs sharing a common algorithmic base. The baseline AAC provides coding bit rates from 16 kbit/s to 256 kbit/s with sampling rates ranging from 7350 Hz to 96 kHz, though the most common rates are 24 kHz, 32 kHz, 44.1 kHz and 48 kHz. It supports mono, stereo, and up to 48 channels of audio including 5.1/7.1-channel surround-sound. While the predominant use of AAC today is for music downloads, a low delay form of AAC (AAC-LD) exists targeted at high-end audio and video conferencing. Therefore, the use of AAC spans two distinct applications, one-way audio, which includes streaming and music downloads, and real-time conferencing.

The 3GPP Release 6 specifications require support of either Extended AMR-WB or Enhanced aacPlus (or both) for packet-switched streaming services (PSS) [ETSI TS 126 234], multimedia messaging services (MMS) [ETSI TS 126 140], and multimedia broadcast/multicast services (MBMS) [ETSI TS 126 346]. For PSS and MMS, 3GPP Release 6 also allows, as an option, use of the MPEG-4 AAC-LC and AAC long-term prediction (AAC-LTP) decoders and encoders. It can be assumed that any Enhanced aacPlus decoder implementation will also be capable of decoding AAC-LC.

8.6.1.1.1 AAC-LC

The simplest form of AAC, AAC low complexity (AAC-LC), is used by Apple iTunes online music store at an aggregate rate of 128 kbit/s (64 kbit/s coding rate for each channel of stereo audio). This provides for audio quality comparable or better than MP3 at 160 kbit/s. (The iTunes application provided by the Mac OS for creating AAC-LC files from CDs allows for user-selectable coding rates.)

---

[1] ™ iPod is a trademark of Apple Inc.

AAC-LC incurs an algorithmic delay of around 100 ms at rates > 80 kbit/s. Delay increases significantly at lower rates to over 300 ms at 24 kbit/s. Consequently, AAC-LC is unsuitable for conferencing applications. However, this does not negate its use in the intended streaming and download applications.

User equipment and media servers supporting any form of AAC for audio streaming SHOULD support AAC-LC for both mono and stereo audio and MAY additionally support AAC-LTP.

8.6.1.1.2 Enhanced aacPlus

The Enhanced aacPlus codec is specified by 3GPP in [ETSI TS 126 401], with the AAC encoder part specified in [ETSI TS 126 403]. Compared with AAC-LC, Enhanced aacPlus allows operation at a lower bit rate by incorporating spectral band replication (SBR) [ETSI TS 126 404] and parametric stereo (PS) [ETSI TS 126 405], which are both defined in [ISO/IEC 14496-3]. The combination of AAC-LC with SBR is referred to as aacPlus by 3GPP and is known as high efficiency AAC (HE-AAC) by MPEG-4. Specifically, within the 3GPP profile, the HE-AAC profile at Level 2 is used as defined in [ISO/IEC 14496-3] and, for stereo terminals, the SBR tool is operated in HQ mode. Enhanced aacPlus, or HE-AAC v2 within MPEG-4, adds parametric stereo in the baseline mode to HE-AAC as defined in clause 8 of [ISO/IEC 14496-3]. The result is a codec that provides for good quality stereo music transmission at rates as low as 24 kbit/s.

The Enhanced aacPlus decoder specified by 3GPP incorporates other tools that are not part of the MPEG-4 audio standard, the most important of which provides for error concealment against frame loss [ETSI TS 126 402]. Error concealment for the AAC core is based on generation of spectrally-shaped noise adapted to the signal while for the SBR and parametric stereo, it is based on extrapolation of guidance, envelope, and stereo information.

Enhanced aacPlus has been extensively tested by 3GPP during its selection and characterization process. These tests included a variety of content types at various bit rates and the evaluation was performed using the MUSHRA methodology [b-MUSHRA]. [b-ETSI TR 126 936] provides the results of this analysis and compares Enhanced aacPlus with the other 3GPP-mandated coder for audio applications (AMR-WB+). The 3GPP PSS [ETSI TS 126 234], MMS [ETSI TS 126 140], MBMS [ETSI TS 126 346] and IMS messaging and presence [b-ETSI TS 126 141] specifications also provide guidance on the audio conditions under which Enhanced aacPlus provides higher performance than AMR-WB+ and vice versa.

User equipment and media servers supporting any form of AAC for audio streaming SHOULD support Enhanced aacPlus for mono and stereo audio. User equipment and media servers supporting enhanced aacPlus decoders MAY support the additional decoder tools specified in [ETSI TS 126 402] for error concealment, stereo-to-mono down-mix, and re-sampling.

8.6.1.1.3 AAC-LD

AAC low delay (AAC-LD) is a derivative of AAC-LC targeted at conferencing applications where round-trip delay is a key performance criterion. AAC-LD has an algorithmic delay of just 20 ms. AAC-LD provides an audio quality that is either comparable with or better than MP3 at the same rate, which is typically in the range from 32 kbit/s to 128 kbit/s per channel although can be as low as 16 kbit/s. Typical sampling rates used with AAC-LD are 32 kHz to provide for a 14 kHz frequency range and 48 kHz for an audio bandwidth of 20 kHz.

To minimize the impact of frame loss, the error resilience tool defined for AAC-LC can also be used with AAC-LD.

AAC-LD is currently not specified by 3GPP as the Release 6 IMS specifications for conversational services do not go beyond wideband. AAC-LD is, however, allowed for conferencing over circuit-switched technology controlled through the H.245 protocol such as 3G-324M and ISDN where commercial video conference systems using the AAC-LD codec already exist. Unlike Enhanced aacPlus, which is specified by 3GPP with bit-exact C-code, no standard code exists for AAC-LD (decoder or encoder).

User equipment and media servers supporting conferencing with super-wideband audio MAY support AAC-LD. User equipment and media servers supporting AAC-LD SHOULD support the error resilience tools specified in [ISO/IEC 14496-3]. User equipment and media servers supporting AAC-LD MUST allow a minimum coding rate during active speech of 48 kbit/s per audio channel with 64 kbit/s RECOMMENDED.

8.6.1.1.4 Packet loss concealment

AAC includes decoder tools to conceal frame loss. For AAC, user equipment and media servers MUST employ the error concealment tool.

8.6.1.1.5 Payload header format

For packet-switched streaming service (PSS) [ETSI TS 126 234], 3GPP specifies the payload format described in [IETF RFC 3016] for Enhanced aacPlus and other forms of AAC. This choice was made mainly for the reason of backwards compatibility with older versions of PSS. For new 3GPP services, which do not have this constraint of backwards compatibility, such as multimedia broadcast/multicast service (MBMS) [ETSI TS 126 346], 3GPP specifies a more flexible payload format described in [IETF RFC 3640]. ISMA [b-ISMA-2] and DVB [ETSI TS 102 005] also mandate the use of [IETF RFC 3640] for the carriage of AAC over IP. Besides broader applicability, [IETF RFC 3640] provides additional features, such as interleaving. User equipment and media servers supporting AAC in streaming services MUST support the payload format specified in [IETF RFC 3016] and MAY support the payload format specified in [IETF RFC 3640]. For any other service using AAC, user equipment and media servers MUST support the payload format specified in [IETF RFC 3640].

[IETF RFC 3016] applies when using the MPEG-4 low-overhead audio transport multiplex (LATM). When used with RTP, LATM allows for the concatenation of multiple audio frames into one stream and the addition of configuration information. LATM streams consist of a sequence of MPEG-4 audioMuxElements that each include one or more audio frames. Where possible, user equipment and media servers SHOULD map one complete audioMuxElement into an RTP packet. When an audioMuxElement exceeds the maximum RTP packet size, user equipment and media servers MAY fragment the audioMuxElement across multiple packets and the RTP Marker bit MUST be set to 0 on all but the last fragment of audioMuxElements.

[IETF RFC 3640] defines an RTP payload structure to transport MPEG-4 elementary streams, of which audio is one type. It can be used for any MPEG-4 codec and can be configured in a flexible way to match the need of individual codecs. Similar to the LATM, multiple MPEG-4 access units may be mapped into one RTP packet but they may not be split across packet boundaries at the same time, i.e., a packet contains either several complete access units or a single fragmented access unit. Mixing complete and fragmented access units in one RTP packet is not allowed, which simplifies implementation and improves interoperability.

The choice of payload format is made through SDP negotiation.

User equipment and media servers SHOULD set the timestamp resolution to the audio sampling rate subject to specifying this via the SDP exchange. If the audio sampling rate is unknown, user equipment and media servers SHOULD set the timestamp resolution to 90 kHz.

8.6.1.1.6 Session description

[IETF RFC 3016] specifies the media encoding name as MP4A-LATM. The information carried in the MIME media type specification has a specific mapping to fields in SDP as follows:

– The media type (audio) goes in SDP "m=" as the media name.
– The media subtype ("MP4A-LATM") goes in SDP "a=rtpmap" as the encoding name.
– The required parameter "rate" also goes in "a=rtpmap" as the clock rate set to the audio sampling rate when this is known or 90000 otherwise.
– The number of channels is not included as a parameter in "a=rtpmap".
– The optional parameter "ptime" goes in the SDP "a=ptime" attribute.
– The optional parameter "profile-level-id" defined by [ISO/IEC 14496-1] may be included to indicate the coder capability in an SDP "a=fmtp" attribute line. The value of 30 (natural audio profile level 1) is the default for when this is not included.
– The "object" parameter defined in MPEG-4 Part 3 [ISO/IEC 14496-3] clause 1.5.1 to specify the flavor of AAC is required in an SDP "a=fmtp" attribute line.
– The payload-format-specific parameters "bitrate", "cpresent", and "config" also go in the "a=fmtp" line. The "config" parameter corresponds with the StreamMuxConfig defined in MPEG-4 Part 3 [ISO/IEC 14496-3] clause 1.7.3.2.3. The parameter "config" is only included when "cpresent" is set to 0; otherwise "cpresent" is set to 1 and the audio configuration information is embedded in the payload. (When the rate is set to 90000, the actual sampling rate may be readable by decoding the config value.)

[IETF RFC 3640] specifies the media encoding name as mpeg4-generic. The information carried in the MIME media type specification has a specific mapping to fields in SDP as follows:

– The media type (audio) goes in SDP "m=" as the media name.
– The media subtype ("mpeg4-generic") goes in SDP "a=rtpmap" as the encoding name.
– The required parameter "rate" goes in "a=rtpmap" as the clock rate set to the audio sampling rate.
– The optional parameter specifying the number of audio channels must also be included in "a=rtpmap" but may be omitted for mono audio streams (i.e., the default is 1).
– The optional parameter "ptime" goes in the SDP "a=ptime" attribute.
– The parameter "streamtype=5" should be included. Streamtype is specified by MPEG-4 Part 1 [ISO/IEC 14496-1] Table 9.
– The optional parameter "profile-level-id" defined by [ISO/IEC 14496-1] may be included to indicate the coder capability in an SDP "a=fmtp" attribute line. The value of 30 (natural audio profile level 1) is the default for when this is not included.
– The parameter "mode=" must be included and set to either "AAC-lbr" (low bit rate) or "AAC-hbr" (high bit rate). The low-bit-rate mode prohibits fragmentation and limits the AAC frame size to 63 octets.
– The "config" parameter corresponds with the AudioSpecificConfig defined in MPEG-4 Part 3 [ISO/IEC 14496-3] clause 1.6.
– The parameters "sizeLength", "indexLength", and indexDeltaLength" must all be included and set as specified by [IETF RFC 3640].

Dynamic payload type numbers are used when declaring or negotiating use of AAC for both payload formats.

8.6.1.2 Extended AMR-WB (AMR-WB+)

Extended AMR-WB (AMR-WB+) [ETSI TS 126 290] is a super-wideband codec that provides high quality performance over a range of audio types at low bit rates. It has been chosen as one of two mandatory codecs by 3GPP for decoders supporting audio in the following applications:

- Packet-switched streaming services (PSS) [ETSI TS 126 234]
- Multimedia messaging services (MMS) [ETSI TS 126 140]
- Multimedia broadcast/multicast services (MBMS) [ETSI TS 126 346]

In addition, AMR-WB+ has been designated an optional codec by ETSI in its generic toolbox for DVB-compliant delivery over RTP and in IP datacast over DVB [ETSI TS 102 005].

AMR-WB+ supports input sampling rates of 8, 16, 24, 32 or 48 kHz, which are then converted to internal sampling frequencies of 12.8 kHz to 38.4 kHz corresponding to audio bandwidths of 6.4 Hz to 19.2 kHz. AMR-WB+ provides coding bit rates of 5.2 to 36 kbit/s for mono signals and 6.2 to 48 kbit/s for stereo.

For the case of mono signals, the input is first converted to the internal sampling frequency and then split into two critically sampled frequency bands for further processing. The low-frequency band can be processed by either of two coding techniques in the 'core' codec. The first possible coding methodology is ACELP employed in a very similar manner to that in AMR-WB [ETSI 126 190]. In fact, AMR-WB+ contains several of the coding modes of AMR-WB. The other choice for the low-frequency band is processing by TCX (transform coded excitation) coding where spectral coefficients are quantized using scalable vector quantization. The choice of ACELP versus TCX can either be made in a closed-loop or an open-loop fashion – with the latter offering a reduced complexity, lower quality variant of the codec. This hybrid ACELP/TCX approach allows AMR-WB+ to deal well with a variety of audio signals as the ACELP encoding is adapted to speech signals and the TCX encoding is designed to handle non-speech audio such as music. As an illustration, the ACELP coding is selected around 1% of the time by AMR-WB+ in dealing with instrumental music input but 48% when encoding speech inputs [b-IEEE COM1]. The high-frequency band is coded using BWE (bandwidth extension) which entails the derivation of a parametric representation of the spectral envelope and temporal gains.

Stereo encoding is achieved by again separating out the low-frequency and high-frequency bands of each channel signal. The two low-frequency components are down-mixed to form a mono signal which is encoded using the standard ACELP/TCX core codec. The two channels are then further split into a very-low-frequency (VLF) band and a mid-band signal. The VLF signal is used to derive a side signal which is encoded in the frequency domain using algebraic VQ. The mid-band signal is parametrically encoded using a shape-gain constrained time-domain filter. The high-frequency bands of both left and right signals are finally encoded separately using the BWE approach. At the decoder, it is possible to down-mix the two channels to provide a mono output.

AMR-WB+ has been extensively tested by 3GPP during its selection and characterization process. These tests included a variety of content types at various bit rates and the evaluation was performed using the MUSHRA methodology [b-MUSHRA]. [b-ETSI TR 126 936] provides the results of this analysis and compares AMR-WB+ with the other 3GPP-mandated coder for audio applications (Enhanced aacPlus). The 3GPP PSS [ETSI TS 126 234], MMS [ETSI TS 126 140], MBMS [ETSI TS 126 346] and IMS messaging and presence [b-ETSI TS 126 141] specifications also provide guidance on the audio conditions under which AMR-WB+ provides higher performance than Enhanced aacPlus and vice versa.

8.6.1.2.1 Packet loss concealment

User equipment and media servers supporting AMR-WB+ MUST use the methods defined in [ETSI TS 126 290] for frame erasure concealment.

8.6.1.2.2 Payload header format

User equipment and media servers MUST support the payload header format as specified in [IETF RFC 4352] for AMR-WB+. In particular, user equipment and media server AMR-WB+ encode implementations MAY support either basic or interleaved mode or both. User equipment and media server AMR-WB+ decode implementations MUST support both basic and interleaved.

8.6.1.2.3 Session description

Parameters are mapped to SDP in a standard way. When conveying information by SDP, user equipment and media servers MUST set the encoding name to "AMR-WB+" (the same as the MIME subtype [IETF RFC 4352]). User equipment and media servers MUST set the RTP clock rate to 72000 and the number of channels MUST be set to 1 or 2, or be omitted, the latter implying a default value of 2.

The following is an example of the media representation in SDP for describing stereo encoding using AMR-WB+ when interleaved mode is employed with a de-interleaving buffer that covers 30 transport frame slots and a minimum of 86400 RTP timestamp ticks:

```
m=audio 49130 RTP/AVP 98
a=rtpmap: 98 AMR-WB+/72000/2
a=maxptime: 100
a=fmtp: 98 interleaving=30; int-delay=86400
```

8.6.1.3 Dolby digital (AC-3)

AC-3 is a high-quality codec supporting multiple audio channels at a low bit-rate [ATSC A/52B]. Exploiting psycho-acoustic phenomena, it achieves substantial data reduction by removing inaudible information from an audio stream. AC-3 supports sampling rates of 32 kHz, 44.1 kHz and 48 kHz, and has data rates ranging from 32 kbit/s to 640 kbit/s, depending on the number of channels and the desired audio quality.

The codec supports configurations up to 5.1 channels, with the following configuration options:

- Mono (Centre only)
- Dual-mono (1 + 1) – two independent mono programs carried in a single bitstream
- 2-channel stereo (Left + Right), optionally carrying matrixed Dolby Surround
- 3-channel stereo (Left, Centre, Right)
- 2-channel stereo with mono surround (Left, Right, Surround)
- 3-channel stereo with mono surround (Left, Centre, Right, Surround)
- 4-channel quadraphonic (Left, Right, Left Surround, Right Surround)
- 5-channel surround (Left, Centre, Right, Left Surround, Right Surround)

All configurations, except for dual mono, can optionally include the extra low frequency effect (LFE) channel (subwoofer).

AC-3 has been adopted by many standard bodies. It is a mandatory audio codec for DVD-Video, ATSC (Advanced Television Systems Committee) digital terrestrial television, DLNA (Digital Living Network Alliance) home networking. It is also an optional multichannel audio codec for DVD-Audio.

AC-3 is also used for streaming applications such as streaming movies from a home media server to a display, video on demand, and multichannel Internet radio.

8.6.1.3.1 Packet loss concealment

When performing fragmentation of an AC-3 frame, user equipment and media servers MAY fragment it such that at least the first 5/8ths of the frame data is in the first fragment. This provides greater resilience to packet loss, since this portion of the frame data is guaranteed to contain the data necessary to decode the first two blocks of the frame.

8.6.1.3.2 Payload header format

User equipment and media servers MUST support the payload header format for AC-3 specified in [IETF RFC 4184]. In constructing an RTP packet, a standard RTP header is used along with a RTP payload, which starts with the two-byte payload header followed by an integral number of complete AC-3 frames or by a single fragment of an AC-3 frame.

8.6.1.3.3 Session description

The information carried in the MIME media type specification has a specific mapping to fields in SDP. When SDP is used to specify sessions employing AC-3, the mapping is as follows:

– The Media type ("audio") goes in SDP "m=" as the media name.
– The Media subtype ("ac3") goes in SDP "a=rtpmap" as the encoding name.
– The required parameter "rate" also goes in "a=rtpmap" as the clock rate, optionally followed by the parameter "channel".
– The optional parameters "ptime" and "maxptime" go in the SDP "a=ptime" and "a=maxptime" attributes, respectively.

An example of the SDP data for AC-3:

```
m=audio 49111 RTP/AVP 100
a=rtpmap:100 ac3/48000/6
```

Certain considerations are needed when SDP is used to perform offer/answer exchanges [IETF RFC 3264].

– The "rate" is a symmetric parameter, and the answer MUST use the same value or remove the payload type. The RTP timestamp clock rate is equal to the audio sampling rate. Permitted rates are 32'000 (32 kHz), 44'100 (44.1 kHz), and 48'000 (48 kHz).
– The "channels" parameter is declarative and indicates, for recvonly or sendrecv, the desired channel configuration to receive, and for sendonly, the intended channel configuration to transmit. This MUST be a number between 1 and 6. The LFE (".1") channel MUST be counted as one channel. All receivers are capable of receiving any of the defined channel configurations, and the parameter exchange might be used to help optimize the transmission to the number of channels the receiver requests. If the "channels" parameter is omitted, a default maximum value of 6 is implied.
– The "ptime" and "maxptime" parameters are negotiated as defined for "ptime" in [IETF RFC 3264].

8.6.1.4 Dolby digital plus (E-AC-3)

E-AC-3 is an enhancement and extension of AC-3 specified in [ATSC A/52B]. It enables operation at both higher and lower data rates than AC-3, and provides expanded channel configurations and the ability to carry multiple programs within a single bitstream through the use of a flexible substream structure. Each substream is equivalent to a conventional AC-3 bitstream, enabling coding of up to 5.1 channels of audio. Up to 8 independent substreams can be present in a bitstream, enabling 8 independent programs to be carried, and each independent substream can have up to 8 dependent substreams associated with it to enable delivery of programs containing more than 5.1 channels (e.g., 7.1 and 13.1). Each substream can be coded at data rates ranging from 32 kbit/s to 6.144 Mbit/s, depending on the number of channels and the desired audio quality. E-AC-3 supports sampling rates of 32 kHz, 44.1 kHz and 48 kHz.

The E-AC-3 specification is published by ETSI as well as the Advanced Television Systems Committee (ATSC). It is an optional codec for ATSC and DVB (digital video broadcasting) television transmission. It is mandatory for use in the High Definition (HD)-DVD-Video optical-storage media format and optional in the Blu-ray Disc format; in both cases, the maximum channel configuration supported is 7.1.

E-AC-3 can also be used for streaming applications such as Internet Protocol television (IPTV), video on demand, interactive features of next generation DVD formats, and transfer of movies across a home network.

8.6.1.4.1 Packet loss concealment

E-AC-3 does not include any inherent packet loss concealment mechanism.

8.6.1.4.2 Payload header format

User equipment and media servers MUST support the payload header format for E-AC-3 specified in [IETF RFC 4598]. In contrasting a RTP packet, a standard RTP header is used along with a RTP payload, which starts with the two-byte payload header followed by an integral number of complete E-AC-3 frames or by a single fragment of an E-AC-3 frame.

8.6.1.4.3 Session description

The information carried in the MIME media type specification has a specific mapping to fields in SDP. When SDP is used to specify sessions employing AC-3, the mapping is as follows:

– The Media type ("audio") goes in SDP "m=" as the media name.
– The Media subtype ("eac3") goes in SDP "a=rtpmap" as the encoding name. The required parameter "rate" also goes in "a=rtpmap" as the clock rate. (The optional "channels" rtpmap encoding parameter is not used. Instead, the information is included in the optional parameter bitStreamConfig.)
– The optional parameters "ptime" and "maxptime" go in the SDP "a=ptime" and "a=maxptime" attributes, respectively.
– The optional parameter "bitStreamConfig" goes in the SDP "a=fmtp" attribute.

The following is an example of the SDP data for E-AC-3 [IETF RFC 4598]:

```
m=audio 49111 RTP/AVP 100
a=rtpmap:100 eac3/48000
a=fmtp:100 bitStreamConfig i6d8d14i6d8
```

Certain considerations are needed when SDP is used to perform offer/answer exchanges [IETF RFC 3264].

– The "rate" is a symmetric parameter, and the answer MUST use the same value or the answerer removes the payload type.
– The "bitStreamConfig" parameter is declarative and indicates, for sendonly, the intended arrangement of substreams in the bit stream, along with the channel configuration, to transmit, and for recvonly or sendrecv, the desired bit stream arrangement and channel configuration to receive. The format of the bitStreamConfig value in an answer MAY differ from the offer value by replacing the number of channels for any undesired substreams with '0'. It is valid to zero out dependent substreams containing undesired channel configurations and to zero out all the substreams of an undesired program. Then the sender MAY reoffer the stream in the receiver's preferred configuration if it is capable of providing that configuration. Note that all receivers are capable of receiving, and all decoders are capable of decoding, any of the legal bit stream configurations, so the parameter exchange is not needed for interoperability. The parameter exchange might be used to help optimize the transmission to the number of programs or channels the receiver requests.

– Since an AC-3 bit stream is a special case of an E-AC-3 bit stream, it is permissible for an AC-3 bit stream to be carried in the E-AC-3 payload format. To ensure interoperability with receivers that support the AC-3 payload format but not the E-AC-3 payload format, a sender that desires to send an AC-3 bit stream in the E-AC-3 payload format SHOULD also offer the session in the AC-3 payload format by including payload types for both media subtypes: 'ac3' and 'eac3'.

8.6.1.5 MP3

MP3 is a common reference to the MPEG-1 Part 3 Layer 3 [ISO/IEC 11172-3] and MPEG-2 Part 3 Layer 3 [ISO/IEC 13818-3] audio codecs. MP3 was originally designed for efficient distribution of music files over moderate bandwidth connections, such as the Internet. As the first mass adopted audio codec for the Internet, there is a large quantity of music now available in the MP3 format.

MPEG-1 Layer 3 works for both mono and stereo signals. It supports sampling rates of 32, 44.1, and 48 kHz. MPEG-2 Layer 3 added multichannel audio coding and lower sampling frequencies of 16, 22.05, and 24 kHz. Collectively, MPEG Layer 3 supports a range of bit-rates from 8 to 320 kbit/s.

8.6.1.5.1 Packet loss concealment

MP3 does not include any inherent packet loss concealment mechanism.

8.6.1.5.2 Payload header format

User equipment and media servers MUST support the payload header format for MP3 specified in [IETF RFC 3555].

8.6.1.5.3 Session description

The information carried in the MIME media type specification has a specific mapping to fields in SDP. When conveying information by SDP, the encoding name MUST be "MPA" (the same as the MIME subtype [IETF RFC 3555]). User equipment and media servers MUST set the RTP clock rate to 90000. The optional "channels" rtpmap encoding parameter is not used. User equipment and media servers MUST include the "layer" parameter in SDP using the "a=fmtp" attribute. User equipment and media servers MUST set the "layer" parameter to "3". When SDP is used to specify sessions employing MP3, the mapping is as follows:

– The Media type ("audio") goes in SDP "m=" as the media name.
– The Media subtype ("mpa") goes in SDP "a=rtpmap" as the encoding name.
– The optional parameters "samplerate", "mode" and "bitrate" go in the SDP "a=fmtp" attribute.
– The optional parameters "ptime" and "maxptime" go in the SDP "a=ptime" and "a=maxptime" attributes, respectively.

An example of the SDP data for MP3:

```
m=audio 49000 RTP/AVP 121
a=rtpmap:121 mpa/90000
a=fmtp: 121 layer=3
```

In case the samplerate and mode parameters are omitted from SDP, they are indicated in the payload.

8.6.2 Codec naming for super-wideband codecs

The super-wideband codecs defined in this specification MUST be encoded by user equipment and media servers with the following string names as defined in Table 8-6 in the rtpmap parameter:

Table 8-6 – Super-wideband audio codec rtpmap parameters

Codec	Literal codec name	rtpmap parameter
AAC (RFC 3016)	MP4A-LATM	MP4A-LATM/24000
AAC (RFC 3640)	AAC	mpeg4-generic/48000/6
Extended AMR-WB	AMR-WB+	AMR-WB+/72000/2
AC-3	AC-3	ac3/48000/6
Enhanced AC-3	E-AC-3	eac3/48000
MP3	MP3	mpa/90000

The timestamp frequency and number of channels following the timestamp frequency are examples. Depending on the codec, the number of channels may be required, is optional or is omitted. When optional, the default number of channels is codec specific.

Unknown rtpmap parameters SHOULD be ignored if they are received by user equipment and media servers.

The bandwidth requirements for super-wideband audio codecs cannot be determined from the media rtpmap attribute (a) lines alone. In this situation, user equipment and media servers MUST use the bandwidth parameter (b) line in SDP to specify its bandwidth requirements. The bandwidth parameter line (b) is of the form:

```
b=<modifier>:<bandwidth>
```

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, the bandwidth parameter should contain the least-upper-bound (LUB) of the desired codec bandwidths.

8.7 Video codec specification

The IPCablecom architecture enables cable operators to provide a wide range of IP multimedia services. In addition to VoIP and data-centric services, video-over-IP represents another important application area for IPCablecom. From the perspective of cable customers, IPCablecom video services can significantly enrich their communication and entertainment experiences, thus strengthening their preference of using the cable network for broadband access and digital entertainment.

IPCablecom video services can potentially be delivered over different hardware platforms, which include the following:

- Standalone video telephony system: This is the traditional video telephony platform. Operating in the IPCablecom environment, this platform can provide increased video quality due to the higher bandwidth and enhanced QoS afforded by the IPCablecom network, with the video quality ultimately being limited by the capability of the integrated display device.
- Wireless CPE: With their mobility, wireless CPE devices (such as 3G handsets, Wi-Fi IP phones and Wi-Fi-enabled PDAs) are emerging as an important video platform fulfilling the niche role for mobile interactive and streaming video applications. In the past, the video quality of a wireless CPE was severely limited by both the bandwidth of the wireless network and the capability of the integrated display device. However, with increasing bandwidth of the wireless network and improving display capability, new classes of wireless CPE devices, supporting ever-increasing video resolutions, are quickly emerging.

• Set-top box (STB): As an asset owned by cable operators, the STB plays a critical role in providing customers with traditional digital TV programming and associated video services (e.g., VOD and PVR). With IPCablecom, the STB's role can be expanded to offer complementary IP-based video telephony and entertainment, reinforcing the prominence of cable operators in providing innovative video services to the customers. These complementary services can greatly benefit from the natural video interface of the television display, especially with its high-definition capability and wide screen.

In addition, by standardizing video codecs, IPCablecom can support interoperability and feature interaction among different video platforms. For instance, a cable customer can use the STB to conduct video conferencing on TV with a remote cellular user. Also, the customer can instruct the STB to re-render the locally stored PVR video content and stream it to a remote cellular user, and vice versa.

As can be seen, an important characteristic of IPCablecom video-over-IP applications is their diversity. In general, such diversity can be viewed from two perspectives or dimensions:

1) Video resolution: Compatible with their video display devices and available network bandwidth, video applications target different video resolutions, such as QCIF (176 × 144), CIF (352 × 288), SD (640 × 480) and HD (1280 × 780 or 1920 × 1080).
2) Video stream direction: Different video applications deal with three different video stream directions: send only (one-way encode), receive only (one-way decode), and simultaneous send and receive (two-way encode/decode).

Since 1990, various video codecs have been developed to cater for different applications, mainly through two international standards bodies – Video Coding Experts Group (VCEG) of ITU-T and Moving Picture Experts Group (MPEG) of ISO/IEC. [b-ITU-T H.261] was developed by ITU-T as the first video codec for video conferencing. MPEG-1 was introduced for video compact disk storage, and was evolved into MPEG-2 (or [ITU-T H.262] as adopted by ITU-T) as the standard codec for DVD, digital television and HDTV. Subsequent H-series and MPEG-series video codecs include [b-ITU-T H.263] and MPEG-4 visual (Part 2) [ISO/IEC 14496-2]. More recently, a high-performance and general-purpose codec, H.264/AVC, has been developed jointly by ITU-T and ISO/IEC.

Associated with each of these video codecs are its profiles and levels. A profile defines a set of coding tools or algorithms that can be used in generating a compliant video bit stream, and a level places constraints on certain key parameters of the bit stream. Each profile/level combination of a codec represents a conformance point that facilitates the interoperability among different video devices. Collectively, profiles and levels define the theoretical capability and flexibility of the associated codec.

As a new-generation architecture for IP multimedia services over the cable network, IPCablecom requires video codecs that are versatile and future-proof. At the same time, it needs to accommodate codecs that are used in legacy video systems. Furthermore, to support cable/cellular integration initiatives, IPCablecom video codec requirements need to be compatible with video codecs standardized by cellular standards bodies such as 3GPP and 3GPP2.

8.7.1 Supported codecs

This clause describes every video codec supported in IPCablecom. Whether a codec is mandatory, recommended or optional depends on the application for which it is used. Therefore, the normative status of each codec is indicated in the associated application capability documents. However, if a particular codec is supported for an application, all the requirements for that codec as specified in this clause MUST be met.

8.7.1.1 H.263

First approved in 1996, [b-ITU-T H.263] is a video codec standardized by ITU-T VCEG [ITU-T G.722] for low bit rate video telephony. It was designed initially for circuit-switched networks such as PSTN, and has since been applied for ISDN and packet networks. H.263 MAY be supported on user equipment and media gateways.

H.263 incorporates improvements over H.261, the previous ITU-T Recommendation for video telephony [b-ITU-T H.261], in the areas of performance and error recovery. It has been designed to stream video at bandwidths as low as 20-24 kbit/s. As a general rule, H.263 improves the coding efficiency over H.261 by 100% (i.e., requires half the bandwidth to achieve the same video quality). In addition, Annex X to [b-ITU-T H.263] supports a wider range of video resolutions than [b-ITU-T H.261] (which only supports QCIF and CIF). As a result, [b-ITU-T H.263] has essentially replaced [b-ITU-T H.261].

8.7.1.1.1 Profile/level requirements

If user equipment and media gateways support [b-ITU-T H.263], the following requirements apply:

- H.263 profile 0 @ level 45 MUST be supported for QCIF applications.
- H.263 profile 3 @ level 45 SHOULD be supported for QCIF applications.
- H.263 profile 0 @ level 40 MUST be supported for CIF applications.
- H.263 profile 3 @ level 40 SHOULD be supported for CIF applications.

The H.263 support for SD and HD application types is not specified for IPCablecom.

The above requirements are summarized in Table 8-7.

Table 8-7 – IPCablecom requirements for H.263

Direction resolution	One-way decode	One-way encode	Two-way codec (interactive)
QCIF	H.263 profile 0 @ level 45 (MANDATORY) H.263 profile 3 @ level 45 (RECOMMENDED)	H.263 profile 0 @ level 45 (MANDATORY) H.263 profile 3 @ level 45 (RECOMMENDED)	H.263 profile 0 @ level 45 (MANDATORY) H.263 profile 3 @ level 45 (RECOMMENDED)
CIF	H.263 profile 0 @ level 40 (MANDATORY) H.263 profile 3 @ level 40 (RECOMMENDED)	H.263 profile 0 @ level 40 (MANDATORY) H.263 profile 3 @ level 40 (RECOMMENDED)	H.263 profile 0 @ level 40 (MANDATORY) H.263 profile 3 @ level 40 (RECOMMENDED)
SD	Not specified	Not specified	Not specified
HD	Not specified	Not specified	Not specified

With the above requirements, user equipment and media gateways supporting QCIF are able to interoperate with mobile devices supporting 3GPP packet-switched conversational multimedia services [ETSI TS 126 235] and 3GPP packet-switched streaming services [ETSI TS 126 234], which all mandate H.263 profile 0 @ level 45 and recommend H.263 profile 3 @ level 45. The same requirements apply to 3GPP2.

H.263 profile and level definitions are given in Annex A.

8.7.1.1.2 Payload header formats

The RTP payload format for H.263-encoded video media MUST be compliant with [IETF RFC 2429] and [IETF RFC 3555].

8.7.1.1.3 Session description

The session description for H.263-encoded video media MUST be compliant with [IETF RFC 2429].

In particular, the MIME media type video/H263-2000 string is mapped to fields in the session description protocol [IETF RFC 4566] as follows:

- The media name in the "m=" line of SDP MUST be 'video'.
- The encoding name in the "a=rtpmap" line of SDP MUST be 'H263-2000' (the MIME subtype).
- The clock rate in the "a=rtpmap" line MUST be '90000'.
- The OPTIONAL parameters "profile" and "level". The "profile" parameter corresponds to the H.263 profile number, in the range 0 through 10, specifying the supported H.263 annexes/subparts. The "level" parameter corresponds to the level of bit stream operation, in the range 0 through 100, specifying the level of computational complexity of the decoding process. The specific values for the "profile" and "level" parameters and their meanings are defined in Annex X of [ITU-T H.263]. Note that the RTP payload format for H.263-2000 is the same as for H.263-1998, but additional annexes/subparts are specified along with the profiles and levels.

An example of media representation in SDP is as follows (profile 0, level 45):

```
m=video 49170 RTP/AVP 98
a=rtpmap:98 H263-2000/90000
a=fmtp:98 profile=0; level=45
b=TIAS:2048000
```

8.7.1.2 H.264/AVC

[b-ITU-T H.264] (or MPEG-4 part 10), is a new generation of video codec jointly developed by ITU-T VCEG and ISO/IEC MPEG. It was first approved in 2003. The codec is also referred to as AVC, for advanced video coding. H.264 MAY be supported on user equipment and media gateways.

H.264/AVC is designed to offer good video quality at bit rates that are substantially lower (e.g., half or less) in comparison with previous video codecs (e.g., MPEG-2, [b-ITU-T H.263], or MPEG-4 part 2). In addition, H.264/AVC is also designed to be a general-purpose codec that can be applied to a wide variety of applications (e.g., low-high bit rates, and low-high video resolutions) and can work robustly on a wide variety of networks and systems (e.g., narrow-band and wideband, wireline and wireless, broadcast, streaming, DVD storage, and video telephony). As reported in one benchmarking [b-ITU-T H.264], H.264/AVC (with main profile) offers a coding-efficiency improvements over MPEG-2 (with main profile), H.263 (with high latency profile) and MPEG-4 (with advanced simple profile) by about 64%, 48% and 38%, respectively. The much improved coding efficiency results from numerous enhancements, including intra-picture prediction, a new 4×4 integer transform, multiple reference frames, variable block sizes, 1/4-pixel precision for motion compensation, a deblocking filter, and enhanced entropy coding.

In general, H.264/AVC has a much higher complexity than the previous video codecs, and requires substantially higher signal-processing capability for encoder and decoder. This is especially the case if full coding efficiency of the codec needs to be realized.

8.7.1.2.1 Profile/level requirements

If user equipment and media gateways support H.264, the following requirements apply:

- H.264 baseline profile @ level 1b MUST be supported for QCIF applications.
- H.264 baseline profile @ level 1.3 MUST be supported for CIF one-way applications.

- H.264 baseline profile @ level 1.2 MUST be supported for CIF two-way applications.
- H.264 baseline profile @ level 1.3 SHOULD be supported for CIF two-way applications.
- H.264 Main profile @ level 3 MUST be supported for SD one-way applications.
- H.264 baseline profile @ level 3 MUST be supported for SD two-way applications.
- H.264 High profile @ level 4 MUST be supported for HD one-way decode applications.

The H.264 support for HD one-way encode and HD two-way encode/decode is not specified for IPCablecom.

When operating in conformance with the baseline profile, an encoder MUST be able to generate a bit stream conformant with constraint_set1_flag = 1, such that the bit stream can be decoded by a main profile decoder. However, if the communicating user equipment or media gateways negotiate the use of any of the baseline profile tools that are not in main profile, e.g., FMO, ASO or redundant slices, the encoders MAY operate with constraint_set1_flag = 0.

These requirements are summarized in Table 8-8.

Table 8-8 – IPCablecom requirements for H.264/AVC

Direction resolution	One-way decode	One-way encode	Two-way codec (interactive)
QCIF	H.264 baseline profile @ level 1b, with constraint_set1_flag = 1 (MANDATORY)	H.264 baseline profile @ level 1b, with constraint_set1_flag = 1 (MANDATORY)	H.264 baseline profile @ level 1b, with constraint_set1_flag = 1 (MANDATORY)
CIF	H.264 baseline profile @ level 1.3, with constraint_set1_flag = 1 (MANDATORY)	H.264 baseline profile @ level 1.3, with constraint_set1_flag = 1 (MANDATORY)	H.264 baseline profile @ level 1.2, with constraint_set1_flag = 1 (MANDATORY) H.264 baseline profile @ level 1.3, with constraint_set1_flag = 1 (RECOMMENDED)
SD	H.264 main profile @ level 3 (MANDATORY)	H.264 main profile @ level 3 (MANDATORY)	H.264 baseline profile @ level 3, with constraint_set1_flag = 1 (MANDATORY)
HD	H.264 high profile @ level 4 (MANDATORY)	Not specified	Not specified

With these requirements, IPCablecom video devices are able to interoperate with 3GPP mobile devices supporting H.264/AVC baseline profile level 1b, which is recommended for 3GPP packet-switched conversational multimedia services [ETSI TS 126 235] and 3GPP packet-switched streaming services [ETSI TS 126 234]. The same requirements apply to 3GPP2.

H.264/AVC profile and level definitions are given in Annex B.

8.7.1.2.2 Payload header formats

The RTP payload format for H.264/AVC-encoded video media MUST be compliant with [IETF RFC 3984] and [IETF RFC 3555].

8.7.1.2.3 Session description

The session description for H.264/AVC-encoded video media MUST be compliant with [IETF RFC 3984] and [IETF RFC 3555].

In particular, the MIME media type video/H264 string is mapped to fields in the session description protocol (SDP) [IETF RFC 4566] as follows:

- The media name in the "m=" line of SDP MUST be 'video'.
- The encoding name in the "a=rtpmap" line of SDP MUST be 'H264' (the MIME subtype).
- The clock rate in the "a=rtpmap" line MUST be '90000'.
- The optional parameters "profile-level-id", "max-mbps", "max-fs", "max-cpb", "max-dpb", "max-br", "redundant-pic-cap", "sprop-parameter-sets", "parameter-add", "packetization-mode", "sprop-interleaving-depth", "deint-buf-cap", "sprop-deint-buf-req", "sprop-init-buf-time", "sprop-max-don-diff" and "max-rcmd-nalu-size", when present, MUST be included in the "a=fmtp" line of SDP. These parameters are expressed as a MIME media type string, in the form of a semicolon separated list of parameter=value pairs.

An example of media representation in SDP is as follows (baseline profile, level 3.0, some of the constraints of the main profile may not be obeyed):

```
m=video 49170 RTP/AVP 98
a=rtpmap:98 H264/90000
a=fmtp:98 profile-level-id=42A01E; sprop-parameter-
sets=Z0IACpZTBYmI,aMljiA==
b=TIAS:10000000
```

8.7.1.3 MPEG-2

MPEG-2 is a video codec standardized by ISO/IEC MPEG [ITU-T H.262] and was first approved in 1995. It has been widely used for traditional cable and satellite digital broadcast TV programming and DVD storage. A large repository of video content is available in MPEG-2-encoded format. MPEG-2 MAY be supported on user equipment and media gateways.

8.7.1.3.1 Profile/level requirements

If user equipment or media gateways support MPEG-2, then the following requirement applies:

- MPEG-2 main profile @ main level SHOULD be supported for SD one-way encode and decode applications.

The MPEG-2 support for other application types is not specified for IPCablecom.

The above requirements are summarized in Table 8-9.

Table 8-9 – IPCablecom requirements for MPEG-2

Direction resolution	One-way decode	One-way encode	Two-way codec (interactive)
QCIF	Not specified	Not specified	Not specified
CIF	Not specified	Not specified	Not specified
SD	MPEG-2 main profile @ main level (RECOMMENDED)	MPEG-2 main profile @ main level (RECOMMENDED)	Not specified
HD	Not specified	Not specified	Not specified

8.7.1.3.2 Payload header formats

The RTP payload format for MPEG-2-encoded video media MUST be compliant with [IETF RFC 2250] and [IETF RFC 3555].

8.7.1.3.3 Session description

The session description for MPEG-2-encoded video media MUST be compliant with [IETF RFC 2250] and [IETF RFC 3555].

In particular, for MPEG-2 transport stream, the MIME media type video/MP2T string is mapped to fields in the session description protocol (SDP) [IETF RFC 4566] as follows:

- The media name in the "m=" line of SDP MUST be 'video'.
- The encoding name in the "a=rtpmap" line of SDP MUST be 'MP2T' (the MIME subtype).
- The clock rate in the "a=rtpmap" line MUST be '90000'.

For MPEG-2 elementary stream, the MIME media type video/MPV string is mapped to fields in the session description protocol (SDP) [IETF RFC 4566] as follows:

- The media name in the "m=" line of SDP MUST be 'video'.
- The encoding name in the "a=rtpmap" line of SDP MUST be 'MPV' (the MIME subtype).
- The clock rate in the "a=rtpmap" line MUST be '90000'.
- The optional parameter "type" MUST be included in the "a=fmtp" line to indicate either "mpeg2-halfd1" (half-D1 video resolution) or "mpeg2-fulld1" (full-D1 video resolution).

An example of MPEG-2 transport stream media representation in SDP is as follows:

```
m=video 49170 RTP/AVP 98
a=rtpmap:98 MP2T/90000
```

An example of MPEG-2 elementary stream media representation in SDP is as follows:

```
m=video 49170 RTP/AVP 98
a=rtpmap:98 MPV/90000
a=fmtp:98 type=mpeg2-fulld1
```

8.7.1.4 MPEG-4 part 2

MPEG-4 part 2 is a video codec that belongs to the MPEG-4 standard family developed by [ISO/IEC 14496-2] and was first approved in 1999. This codec has been employed by some portable devices such as 3G handsets and digital still cameras that capture and playback video clips. As benchmarked in [ISO/IEC 14496-2] MPEG-4 part 2 (with advanced simple profile) offers a coding-efficiency improvement over MPEG-2 (with main profile) and H.263 (with high latency profile) of approximately 42% and 16%, respectively. MPEG-4 part 2 MAY be supported on user equipment and media gateways.

8.7.1.4.1 Profile/level requirements

If user equipment or media gateways support MPEG-4 Part 2, then the following requirements apply:

- MPEG-4 Part 2 simple profile @ level 0b SHOULD be supported for QCIF applications.
- MPEG-4 Part 2 simple profile @ level 2 SHOULD be supported for CIF applications.

The MPEG-2 support for other application types is not specified for IPCablecom.

The above requirements are summarized in Table 8-10.

Table 8-10 – IPCablecom requirements for MPEG-4 part 2

Direction resolution	One-way decode	One-way encode	Two-way codec (interactive)
QCIF	MPEG-4 part 2 simple profile @ level 0b (RECOMMENDED)	MPEG-4 part 2 simple profile @ level 0b (RECOMMENDED)	MPEG-4 part 2 simple profile @ level 0b (RECOMMENDED)
CIF	MPEG-4 part 2 simple profile @ level 2 (RECOMMENDED)	MPEG-4 part 2 simple profile @ level 2 (RECOMMENDED)	MPEG-4 part 2 simple profile @ level 2 (RECOMMENDED)
SD	Not specified	Not specified	Not specified
HD	Not specified	Not specified	Not specified

With the above requirements, IPCablecom video devices are able to interoperate with 3GPP mobile devices supporting MPEG-4 part 2 simple profile level 0b, which is recommended for 3GPP packet-switched conversational multimedia services [ETSI TS 126 235] and 3GPP packet-switched streaming services [ETSI TS 126 234]. The same requirements apply to 3GPP2.

8.7.1.4.2 Payload header formats

The RTP payload format for MPEG-4 part 2-encoded video media MUST be compliant with [IETF RFC 3016] and [IETF RFC 3555].

8.7.1.4.3 Session description

The session description for MPEG-4 part 2-encoded video media MUST be compliant with [IETF RFC 3016] and [IETF RFC 3555].

In particular, the MIME media type video/MP4V-ES string is mapped to fields in the session description protocol (SDP) [IETF RFC 4566] as follows:

- The media name in the "m=" line of SDP MUST be 'video'.
- The encoding name in the "a=rtpmap" line of SDP MUST be 'MP4V-ES' (the MIME subtype).
- The optional parameter "rate" goes in "a=rtpmap" as the clock rate.
- The optional parameter "profile-level-id" and "config" go in the "a=fmtp" line to indicate the coder capability and configuration, respectively. These parameters are expressed as a MIME media type string, in the form of a semicolon-separated list of parameter=value pairs.

The following is an example of media representation in SDP, with simple profile/level 2, rate=90000 (90 kHz), "profile-level-id" and "config" present in "a=fmtp" line:

```
m=video 49170/2 RTP/AVP 98
a=rtpmap:98 MP4V-ES/90000
a=fmtp:98 profile-level-
id=2;config=000001B001000001B50900000100000001200084440FA282C2090A21F
```

8.7.2 Summary of supported codecs

Table 8-11 summarizes all video codecs that are supported by IPCablecom, with profile and level requirements specified with respect to various application types. The choice of which resolution/direction combinations to support is vendor-specific, but if user equipment and media gateways support a particular resolution/direction combination, the requirements specified for that combination MUST be met.

Table 8-11 – Summary of IPCablecom video codec requirements

Direction resolution	One-way decode	One-way encode	Two-way codec (interactive)
QCIF	H.263 profile 0 @ level 45 (MANDATORY) H.264 baseline profile @ level 1b, with constraint_set1_flag = 1 (MANDATORY) H.263 profile 3 @ level 45 (RECOMMENDED) MPEG-4 part 2 simple profile @ level 0b (RECOMMENDED)	H.263 profile 0 @ level 45 (MANDATORY) H.264 baseline profile @ level 1b, with constraint_set1_flag = 1 (MANDATORY) H.263 profile 3 @ level 45 (RECOMMENDED) MPEG-4 part 2 simple profile @ level 0b (RECOMMENDED)	H.263 profile 0 @ level 45 (MANDATORY) H.264 baseline profile @ level 1b, with constraint_set1_flag = 1 (MANDATORY) H.263 profile 3 @ level 45 (RECOMMENDED) MPEG-4 part 2 simple profile @ level 0b (RECOMMENDED)
CIF	H.263 profile 0 @ level 40 (MANDATORY) H.264 baseline profile @ level 1.3, with constraint_set1_flag = 1 (MANDATORY) H.263 profile 3 @ level 40 (RECOMMENDED) MPEG-4 part 2 simple profile @ level 2 (RECOMMENDED)	H.263 profile 0 @ level 40 (MANDATORY) H.264 baseline profile @ level 1.3, with constraint_set1_flag = 1 (MANDATORY) H.263 profile 3 @ level 40 (RECOMMENDED) MPEG-4 part 2 simple profile @ level 2 (RECOMMENDED)	H.263 profile 0 @ level 40 (MANDATORY) H.264 baseline profile @ level 1.2, with constraint_set1_flag = 1 (MANDATORY) H.263 profile 3 @ level 40 (RECOMMENDED) H.264 baseline profile @ level 1.3, with constraint_set1_flag = 1 (RECOMMENDED) MPEG-4 part 2 simple profile @ level 2 (RECOMMENDED)
SD	H.264 Main profile @ level 3 (MANDATORY) MPEG-2 main profile @ main level (RECOMMENDED)	H.264 Main profile @ level 3 (MANDATORY) MPEG-2 main profile @ main level (RECOMMENDED)	H.264 baseline profile @ level 3, with constraint_set1_flag = 1 (MANDATORY) MPEG-2 main profile @ main level (RECOMMENDED)
HD	H.264 high profile @ level 4 (MANDATORY)	Not specified	Not specified

8.7.3 Error recovery

Communication errors can degrade video quality. There are many sources of such errors, including burst bit errors resulting from communication channel impairments and the loss of packets resulting from undesirable network conditions such as congestion.

There exist multiple mechanisms to mitigate the effect of communication errors on video quality:

- forward error correction (FEC);
- packet retransmission;
- error concealment;
- error-resilient coding.

The first two types of error-control mechanisms are application-specific, and are outside of the scope of this Recommendation.

Different video codecs usually have their own algorithms for handling error concealment and error-resilience. These algorithms may also be specific to a particular profile/constraint of a video codec. This Recommendation does not specify any error-concealment and error-resilience algorithms which are not included in the mandated or recommended video codecs.

8.7.4 Codec naming and flowspec parameters for video codecs

The video codecs defined in this Recommendation MUST be encoded with the string names in the "rtpmap" parameters as shown in Table 8-12.

Table 8-12 – Video codecs "rtpmap" parameters

Codec	Literal codec name	"rtpmap" parameter
H.263	H263-2000	H263-2000/90000
H.264/AVC	H264	H264/90000
MPEG-2 transport stream	MP2T	MP2T/90000
MPEG-2 elementary stream	MPV	MPV/90000
MPEG-4 part 2	MP4V-ES	MP4V-ES/90000

Unknown "rtpmap" parameters SHOULD be ignored if they are received.

For every recommended codec (whether it is represented in SDP as a static or dynamic payload type), Table 8-13 describes the mapping that MAY be used from either the payload type or ASCII string representation to the bandwidth requirements for that codec, with the bandwidth requirements being expressed as flowspec.

Table 8-13 – Mapping of video codec session description parameters to flowspec

Parameters from session description			Flowspec parameters		Comments
RTP/AVP code	rtpmap	ptime (ms) (Note 1)	Values b, m, M (Notes 2 and 4)	Values r, p (Notes 3 and 5)	
H.263	H263-2000/90000	N/A	b, M = 1500 bytes m = 128 bytes	Level 45: 20 kbyte/s Level 40: 308 kbyte/s	b, m and M are default values r, p = max compressed bit rate × 120%
H.264	H264/90000	N/A	b, M = 1500 bytes m = 128 bytes	Level 1b: 20 kbyte/s Level 1.2: 58 kbyte/s Level 1.3: 116 kbyte/s Level 3: 1.5 Mbyte/s Level 4: 3 Mbyte/s	b, m and M are default values r, p = max compressed bit rate × 120%
MPEG-2	Transport stream: MP2T/90000 Elementary stream: MPV/90000	N/A	b, M = 1500 bytes m = 128 bytes	Main level: 2.25 Mbyte/s	b, m and M are default values r, p = max compressed bit rate × 120%
MPEG-4 part 2	MP4V-ES/90000	N/A	b, M = 1500 bytes m = 128 bytes	Level 0b: 9.6 kbyte/s Level 2: 19.2 kbyte/s	b, m and M are default values r, p = max compressed bit rate × 120%

Table 8-13 – Mapping of video codec session description parameters to flowspec

NOTE 1 – "Ptime" is not applicable to video in general. An application can set this parameter to an accurate value if it is known.
NOTE 2 – The parameters 'b', 'm' and 'M' are set to their default values, since the packet sizes and maximum burst rates are not regular for video. An application can set these parameters to their accurate values if such values are known.
NOTE 3 – Since the packet rates for video are not regular, the IP/UDP/RTP overhead data rate cannot be derived accurately in general. As a work-around, 20% of the maximum compressed data rate is assumed for the overhead. An application can set these parameters to their accurate values if such values are known.
NOTE 4 – *b* is bucket depth (bytes). '*m*' is minimum policed unit (bytes). '*M*' is maximum datagram size (bytes).
NOTE 5 – *r* is bucket rate (bytes/s). '*p*' is peak rate (bytes/s).

For less 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, the bandwidth parameter should contain the least-upper-bound (LUB) of the desired codec bandwidths.

8.8 Media quality measurement and monitoring

One of the principal goals of IPCablecom is to enhance the user's experience with video and high-fidelity audio. Therefore, it is important for IPCablecom UE to be able to monitor the quality of the audio and video streams. This clause specifies the associated requirements.

Media quality metrics can be characterized along two dimensions: objective vs subjective and intrusive vs non-intrusive. Objective metrics (e.g., PESQ or PSNR) can be computed "on the fly" while the system under test is in service, whereas subjective metrics (e.g., mean opinion scores) are the result of test sessions where a number of people are asked to watch "standard" test clips and to rate their quality. Separately, intrusive (or double-ended) measurement relies on the availability of a reference stream; specifically, a signal is passed through the system under test, and the degraded output is compared with the input (reference) signal. Non-intrusive testing does not rely on such a reference stream. This Recommendation focuses on objective metrics that can be obtained via non-intrusive testing.

8.8.1 Audio quality measurement and monitoring

8.8.1.1 RTCP XR VoIP metric 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 user equipment or media gateways; however, they may be captured by intermediate network probes or analysers, or potentially by embedded monitoring functionality in CMTS and routers. User equipment and media gateways MAY support RTCP XR VoIP metrics.

User equipment and media gateways that support RTCP XR VoIP metrics 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 [IETF RFC 3550].

User equipment and media gateways that support the RTCP XR VoIP Metrics payload MUST measure or compute the reported values of the metrics as defined in [IETF RFC 3611] and clarified in clauses 8.7.1.2 to 8.7.1.7.

8.8.1.2 Reporting of RTCP XR VoIP metrics via SIP

The reporting of RTCP XR VoIP metrics from user equipment or media gateways to a performance management function located in a back-office server is governed by the SIP standard for reporting service quality [ANSI T1.508]. The back-office server collector function that receives the VoIP metrics reports is referred to as the "collector" device in [ID SIP RTCP]. This is typically an element manager or network manager that is responsible for VoIP session/media performance management. During registration, the user equipment and media gateways MUST indicate support of the vq-rtcpxr package defined in [ID SIP RTCP]. The user equipment and media gateways are informed of the contact address of the collector as part of the registration process.

Although [ID SIP RTCP] permits the use of either the PUBLISH or SUBSCRIBE/NOTIFY methods for reporting VoIP metrics, user equipment and media gateways that support RTCP XR VoIP metrics MUST support the use of PUBLISH for this purpose. [ID SIP RTCP] specifies three types of metrics reports: session reports; interval reports; and alert reports. User equipment and media gateways that support RTCP XR VoIP metrics MUST support session reports and MAY support interval reports and alert reports.

User equipment and media gateways that support RTCP XR VoIP metrics MUST send a session report at the end of a session and when the session is reconfigured in any way. Examples of session reconfigurations are call transfers, conference joins, codec changes and changes from one media type (e.g., voice) to another (e.g., voiceband data). User equipment and media gateways that support RTCP XR VoIP metrics MUST cover the time from session start or from the last session change when a report was issued for session reports. User equipment and media gateways that support RTCP XR VoIP metrics MUST include:

- all parameters such as start and stop time stamps, local and remote addresses etc., that are mandatory in [ID SIP RTCP];
- all local metrics and remote metrics listed in [ID SIP RTCP] that are derived from [IETF RFC 3611];
- the local and remote versions of inter-arrival jitter, based on [IETF RFC 3550].

All other parameters and metrics in [ITU-T G.114] are optional.

8.8.1.3 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 as shown in Table 8-14. 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% 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.

Table 8-14 – Metrics related to packet loss and discard

Metric	Description	Range
Loss rate	Proportion of packets lost within the network	0-0.996
Discard rate	Proportion of packets discarded due to late arrival	0-0.996
Burst loss density	Proportion of packets lost and discarded during burst periods	0-0.996
Gap loss density	Proportion of packets lost and discarded during gap periods	0-0.996
Burst duration	Average length of burst periods (ms)	0-65'535
Gap duration	Average length of gap periods (ms)	0-65'535
Gmin	Parameter used to define burst periods	0-255

User equipment and media gateways when using RTCP XR MUST provide these parameters as defined in [IETF RFC 3611].

8.8.1.4 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 an 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.

Table 8-15 – Metrics related to delay

Metric	Description	Range
Round trip delay	Packet path round trip delay (ms)	0-65'535
End system delay	Round trip delay within end system (ms)	0-65'535

User equipment and media gateways using RTCP XR MUST provide the parameters as defined in [IETF RFC 3611]. Note this requires an SR or RR exchange prior to the inclusion of an XR payload into an RTCP message.

8.8.1.5 Definition of metrics related to signal

The signal level, noise level and residual echo return loss are intended to support the diagnosis of problems related to a 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 user equipment or media gateways.

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, in general, this will not represent an accurate measurement of the residual echo but can provide a useful indication of the presence of echo problems. Echo occurring on the endpoint reporting this metric will be heard by the user at the remote endpoint if significant delay is present on the call.

Table 8-16 – Metrics due to signal

Metric	Description	Range
Signal level	RMS signal level during active speech periods (dBm0) as defined in [b-ITU-T P.56] and [b-ITU-T P.561]	–30 to +3
Noise level	RMS noise level during silence periods (dBm0) as defined in [b-ITU-T P.56] and [b-ITU-T 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 [ITU-T G.168]	0 to 80

User equipment and media gateways using RTCP XR MUST provide signal and noise levels as defined in [ITU-T G.107].

An IPCablecom endpoint equipped with an echo canceller, and when using RTCP XR, MUST provide the residual echo return loss metric as defined in [IETF RFC 3611].

8.8.1.6 Definition of metrics related to call quality

Call quality metrics are useful when assessing the overall quality of a call [IETF RFC 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 [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 E-model is a conversational quality metric; however, it can be used to estimate a conversational and listening quality MOS. 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*" reflects 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; note however that this produces a MOS slightly higher than those typically reported from subjective tests.

Table 8-17 – Metrics related to call quality

Metric	Description	Range
R factor	Conversational transmission quality rating	0-100
External *R* factor	*R* factor for an attached external network	0-100
MOS-LQ	Estimated listening quality MOS (×10)	10-50
MOS-CQ	Estimated conversational quality MOS (×10)	10-50

User equipment and media gateways using RTCP XR MUST provide the *R* factor, MOS-LQ and MOS-CQ metrics and MAY provide an external *R* factor.

User equipment and media gateways using RTCP XR MUST calculate *R* factors using G.107 at a minimum [ITU-T G.107].

User equipment and media gateways 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 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 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 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].

User equipment and media gateways using RTCP XR MUST calculate the "Ie-eff" parameter using the function defined in [ITU-T G.107]. However, user equipment and media gateways MUST use the "Ie" and "Bpl" parameters defined in Table 8-18 for the codec and PLC combinations listed.

Table 8-18 – Ie and Bpl parameters for IPCablecom codecs

Vocoder	Bit rate	PLC	Ideal R	Ideal MOS	Ie	Bpl
G.711 A/U	64 kbit/s	Appendix I	93	4.4	0	34
G.728 10 ms	16 kbit/s	Per Annex I to [ITU-T G.728]	89	4.1	7	17
G.728 20 ms	16 kbit/s	Per Annex I to [ITU-T G.728]	89	4.1	7	15
G.729 Annex E 10 ms	11.8 kbit/s	Per [b-ITU-T G.729]	88	4.1	4	20
G.729 Annex E 20 ms	11.8 kbit/s	Per [b-ITU-T G.729]	88	4.1	4	19
ILBC 20 ms	15.2 kbit/s	Per [IETF RFC 3951]	80	3.9	10	34
ILBC 30 ms	13.3 kbit/s	Per [IETF RFC 3951]	78	3.8	12	27
BV16 10 ms	16 kbit/s	Per [IETF RFC 3951]	88	4.2	5	25
BV16 20 ms	16 kbit/s	Per [IETF RFC 3951]	88	4.2	5	23

User equipment and media gateways 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).

User equipment and media gateways 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 as follows:

a) Use [b-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 [b-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 zero 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.

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

$$\text{Estimated PESQ score} = 1 + 0.033.R_{adj} + R_{adj} (100 – R_{adj}) (R_{adj} – 60) \times 0.000007$$

c) Determine the Bpl value using the objective test scores for other packet loss rates. This may be obtained by iteratively searching for the Bpl value that, when converted to an R factor and then an estimated P.862 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 (derived per [ITU-T G.107]) with available absolute category rating (ACR) test data (if available) in order to verify values.

8.8.1.7 Definition of parameters related to endpoint configuration

The parameters in Table 8-19 describe some key configuration parameters of the IPCablecom endpoint that are useful in monitoring service quality and identifying some types of configuration related problems.

User equipment and media gateways using RTCP XR MUST provide values to all parameters as defined in Table 8-19.

Table 8-19 – 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	0-15
Jitter buffer – nominal delay	Nominal delay applied to received packets by the jitter buffer for packets arriving on time	0-65'535
Jitter buffer – maximum delay	Maximum delay applied to received packets by the jitter buffer	0-65'535
Jitter buffer – absolute max delay	Maximum delay size that an adaptive jitter buffer can reach	0-65'535

8.8.2 Video quality and RTCP XR

Recognizing the importance of leveraging the same technology for video as for voice, the IETF AVT working group is in the process of developing a video metrics block for RTCP XR – analogous to the existing voice metrics block. Due to the immaturity of this effort, specification of a video quality metrics block is out of scope for this release of this Recommendation.

Annex A

H.263 profiles and levels

(This annex forms and integral part of this Recommendation)

Tables A.1 and A.2 summarize [b-ITU-T H.263] profiles and levels (Annex X to [b-ITU-T H.263]).

Table A.1 – Summary of H.263 profiles

Annex/clause below for profile listed at right	0	1	2	3	4	5	6	7	8
5.1.5: *Custom picture format (CPFMT)*	L	L	L	L	L	L	L	L	L
5.1.7: *Custom picture clock frequency code (CPCFC)*	L	L	L	L	L	L	L	L	L
C: *Continuous presence multipoint and video multiplexing*									
D.1: *Motion vectors over picture boundaries*		X	X	X	X	X	X	X	X
D.2 with UUI = '1' or UUI not present: *Extension of the motion vector range*						X	X	X	X
D.2 with UUI = '01': *Unlimited extension of the motion vector range*									
E: *Syntax-based arithmetic coding*									
F.2: *Four motion vectors per macroblock*		X	X	X	X	X	X	X	X
F.3: *Overlapped block motion compensation*			X			X	X	X	X
G: *PB-frames*									
H: *Forward error correction (use may be imposed at system level as in ITU-T Rec. H.320)*									
I: *Advanced intra coding*		X		X	X	X	X	X	X
J: *Deblocking filter*		X		X	X	X	X	X	X
K without submodes: *Slice structured coding – Without submodes*				X	X		X		X
K with ASO: *Slice structured coding – With arbitrary slice ordering submode*							X		X
K with RS: *Slice structured coding – With rectangular slice submode*									
L.4: *Supplemental enhancement full picture freeze*		X				X	X	X	X
L: *Supplemental enhancement – Other SEI features*									
M: *Improved PB-frames*									
N: *Reference picture selection (and submodes)*									
O.1.1 Temporal (B pictures): *temporal, SNR, and spatial scalability – B pictures for temporal scalability*									X
O SNR and spatial: *Temporal, SNR, and spatial scalability – EI and EP pictures for SNR and spatial scalability*									
P.5: *Reference picture resampling – Implicit factor of four*									X
P: *Reference picture resampling – More general resampling*									
Q: *Reduced resolution update*									
R: *Independent segment decoding*									

Table A.1 – Summary of H.263 profiles

Annex/clause below for profile listed at right	0	1	2	3	4	5	6	7	8
S: *Alternative inter VLC*									
T: *Modified quantization*		X		X	X	X	X	X	X
U without submodes: *Enhanced reference picture selection – Without submodes*						X	X	X	X
U with SPR: *Enhanced reference picture selection – With sub-picture removal submode*									
U with BTPSM: *Enhanced reference picture selection – With B-picture two-picture submode*									
V: *Data partitioned slices*					X				
W.6.3.8: *Additional SEI specification – Prior picture header repetition*					X				
W.6.3.11: *Additional SEI specification – Interlaced field indications*								X	
W: *Additional SEI specification – Other SEI features*									
"X" indicates that support of a feature is part of a profile. "L" indicates that the inclusion of a feature depends on the level within the profile.									

Table A.2 – Summary of H.263 levels

Level	**10**	**20**	**30**	**40**
Max picture format	QCIF (176 × 144)	CIF (352 × 288)	CIF (352 × 288)	CIF (352 × 288)
Min picture interval	2002/(30'000) s	2002/(30'000) s for CIF 1001/(30'000) s for QCIF and sub-QCIF	1001/(30'000) s	1001/(30'000) s
Max bit rate in 64'000 bit/s units	1	2	6	32
Level	**45**	**50**	**60**	**70**
Max picture format	QCIF (176 × 144) support of CPFMT in profiles other than 0 and 2	CIF (352 × 288) support of CPFMT	CPFMT: 720 × 288 support of CPFMT	CPFMT: 720 × 576 support of CPFMT
Min picture interval	2002/(30'000) s support of CPCFC in profiles other than 0 and 2	1/50 s at CIF or lower 1001/(60'000) s at 352 × 240 or smaller support of CPCFC	1/50 s at 720 × 288 or lower 1001/(60'000) s at 720 × 240 or smaller support of CPCFC	1/50 s at 720 × 576 or lower 1001/(60'000) s at 720 × 480 or smaller support of CPCFC
Max bit rate in 64'000 bit/s units	2	64	128	256

Annex B

H.264/AVC profiles and levels

(This annex forms and integral part of this Recommendation)

Tables B.1, B.2 and B.3 summarize H.264 profiles and levels [b-ITU-T H.264]:

Table B.1 – H.264/AVC original profiles

Coding Tools	Baseline	Main	Extended
I and P slices	X	X	X
CABAC		X	
B slices		X	X
Interlaced coding		X	X
Enhanced error resilience (FMO, ASO, RS)	X		X
Further enhanced error resilience (DP)			X
SP and SI slices			X
I and P slices		X	X

Table B.2 – H.264/AVC New profiles in FRExt amendment

Coding Tools	High	High 10	High 4:2:2	High 4:4:4
Main profile tools	X	X	X	X
4:2:0 chroma format	X	X	X	X
8-bit sample bit depth	X	X	X	X
8×8 vs 4×4 transform adaptivity	X	X	X	X
Quantization scaling matrices	X	X	X	X
Separate Cb and Cr QP control	X	X	X	X
Monochrome video format	X	X	X	X
9- and 10-bit sample bit depth		X	X	X
4:2:2 chroma format			X	X
11- and 12-bit sample bit depth				X
4:4:4 chroma format				X
Residual colour transform				X
Predictive lossless coding				X

Table B.3 – H.264/AVC levels

Level number	Typical picture size	Typical frame rate	Maximum compressed bit rate (for VCL) in non-FRExt profiles	Maximum number of reference frames for typical picture size
1	QCIF	15	64 kbit/s	4
1b	QCIF	15	128 kbit/s	4
1.1	CIF or QCIF	7.5 (CIF)/30 (QCIF)	192 kbit/s	2 (CIF)/9 (QCIF)
1.2	CIF	15	384 kbit/s	6
1.3	CIF	30	768 kbit/s	6
2	CIF	30	2 Mbit/s	6
2.1	HHR (480i or 576i)	30/25	4 Mbit/s	6
2.2	SD	15	4 Mbit/s	5
3	SD	30/25	10 Mbit/s	5
3.1	1280 × 720p	30	14 Mbit/s	5
3.2	1280 × 720p	60	20 Mbit/s	4
4	HD (720p or 1080i)	60p/30i	20 Mbit/s	4
4.1	HD (720p or 1080i)	60p/30i	50 Mbit/s	4
4.2	1920 × 1080p	60p	50 Mbit/s	4
5	2K × 1K	72	135 Mbit/s	5
5.1	2K × 1K or 4K × 2K	120/30	240 Mbit/s	5

Appendix I

Characteristics of narrow-band codecs

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

The information provided in this appendix has been gathered by several people from a variety of sources. As such, it is possible that errors have been made. It is provided for general information. Please consult the appropriate source documents before using it for developmental purposes.

Table I.1 – Narrow-band codecs (part 1)

Codec	Technology	Year standardized	Coding rate (kbit/s)	Codec frame size (ms)	Look ahead (ms)	Algorithmic delay (ms) (Note 1)	Total codec delay (ms) (Note 2)	RAM (kwords) (Note 3)	ROM (kwords)
G.711	Companded PCM	1972	64	0.125	0	0.125	0.25	~0.01	~0.5
G.726	Adaptive Differential PCM	1990	16, 24, 32, 40	0.125	0	0.125	0.25	~0.15	< 2
G.728	LD-CELP	1992	16	0.625	0	0.625	1.25	~2.2	6.7
G.729	CS-ACELP	1995	8	10	5	15	25	~2.6	~14
G.729A	CS-ACELP	1996	8	10	5	15	25	~2.6	~12
G.729E	CS-ACELP	1998	11.8	10	5	15	25	~2.6	~20
G.723.1	MPC-MLQ; ACELP	1995	6.3 and 5.3	30	7.5	37.5	67.5	~2.1	~20
iLBC	FB-LPC	2002	15.2 and 13.3	20 and 30	5 and 10	25 and 40	45 and 70	~4	~12
BV16	TSNFC (2-stage noise feedback coding)	2003	16	5	0	5	10	~2	~11
GSM EFR	ACELP	1995	12.2	20	0	20	40	~4.6	
TDMA IS-641	ACELP	1995	7.4	20	5	25	45	~2.5	
EVRC IS-127	RCELP	1997	0.8, 2.0, 4.0, 8.55	20	10	30	50	~2.5	
IS-733	CELP	1997	1.0, 2.7, 6.2, 13.3	20	5	25	45	~2.5	
AMR	ACELP	1999-2001	4.75, 5.15, 5.9, 6.7, 7.4, 7.95, 10.2, 12.2	20	5	25	45	~4.6	17
IS-893 SMV (Note 4)	eX-CELP	2001	0.8, 2.0, 4.0, 8.5	20	12.5	32.5	52.5	~7.5	

Table I.1 – Narrow-band codecs (part 2)

Codec	Comp-lexity (MIPS) (Note 5)	Codec impairment (G.107 Ie)	Calculated MOS CQE according to G.107 (Note 6)	MOS CQE for intra-MSO calls (Note 7)	MOS CQE for intra-MSO calls (Note 8)	MOS CQE for cable-to-PSTN calls (Note 9)	MOS CQE for cable-to-cellular calls (Note 10)	Known comparison with references in official 3rd party listening MOS tests (Note 11)	Packet loss rate for 0.5 MOS degradation (Note 12)
G.711	~0.35	0	4.41	4.41	4.41	4.41	4.41	Reference	3 % (with App. I)
G.726	~12	50, 25, 7, 2	2.22, 3.51, 4.24, 4.37	2.22, 3.51, 4.24, 4.37	2.22, 3.51, 4.23, 4.37	2.22, 3.51, 4.24, 4.37	2.22, 3.51, 4.23, 4.37	< G.711	No PLC mechanism
G.728	~36	7	4.24	4.24	4.23	4.24	4.23	~ G.726 (32K)	3 %
G.729	~22	10	4.14	4.14	4.12	4.14	4.11	~ G.726 (32K)	3 %
G.729A	~13	11	4.10	4.10	4.09	4.10	4.08	< G.726 (32K)	3 %
G.729E	~27	4	4.32	4.32	4.30	4.32	4.30	= G.726 (32K)	3 %
G.723.1	~19	15 and 19	3.95 and 3.79	3.94 and 3.77	3.80 and 3.61	3.95 and 3.79	3.77 and 3.59	<< G.726 (32K)	3 %
iLBC	~15 and ~18	10 and 12	4.14 and 4.07	4.14 and 4.05	4.08 and 3.91	4.14 and 4.07	4.06 and 3.89	~ G.729E > G.728	7 % and 5 %
BV16	~12	5	4.29	4.29	4.29	4.29	4.29	> G.726 (32K) > G.728 > G.729	5 %
GSM EFR	~18	5	4.29	4.29	4.26	4.29	4.24	~ G.726 (32K)	3 %
TDMA IS-641	~15	10	4.14	4.14	4.08	4.14	4.06	~ G.729	3 %
EVRC IS-127	~25	6	4.27	4.26	4.20	4.27	4.19	< G.726 (32K)	3 %
IS-733	~22							< G.726 (32K)	3 %
AMR	~20	5 (at 12.2 kbit/s)	4.29 (at 12.2 kbit/s)	4.29 (at 12.2 kbit/s)	4.24 (at 12.2 kbit/s)	4.29 (at 12.2 kbit/s)	4.23 (at 12.2 kbit/s)	~ G.726 (32K)	3 %
IS-893 SMV (Note 4)	~40 WMOPS							= G.711 (with mode_0) > EVRC (with mode_0) ~ G.711 (with mode_1)	3 %

Table I.1 – Narrow-band codecs (part 3)

Codec	Performance for background noise	MOS reduction for tandem encodings (Note 13)	Performance	Other functionality	Prevalence	Applications	Status
G.711	Toll quality	Toll quality	Very good for speech, audio, DTMF, text, fax and voiceband data	PLC defined in App. I, SID defined in App. II	High	Ubiquitous	ITU-T
G.726	Toll quality	Toll quality	Good for speech, audio and DTMF; very good for text; poor for fax and modem		High	International links, DCME	ITU-T
G.728	Toll quality	Toll quality	Good for speech		Medium	DCME, video conferencing, IPCablecom E-MTA	ITU-T
G.729	≤ Toll quality	< Toll quality	Good for speech without transcoding	Integrated VAD and PLC	High	IP phones	ITU-T
G.729A	≤ Toll quality	< Toll quality	Good for speech without transcoding	Integrated VAD and PLC	High	IP phones, DSVD	ITU-T
G.729E	Toll quality	Toll quality	Good for speech	Integrated VAD and PLC; music detection	Low	IPCablecom E-MTA	ITU-T
G.723.1	≤ Toll quality	< Toll quality	Acceptable for speech at 6.3 kbit/s; poor at 5.3 kbit/s		Medium	Videophone over dial-up, IP phones	ITU-T but used in enterprise networks
iLBC	Toll quality	< Toll quality	Good for speech without transcoding	Integrated VAD and PLC	Medium	IPCablecom E-MTA, Skype	IETF; IPCablecom E-MTA
BV16	Toll quality	Toll quality	Good for speech	Example PLC available	Low	IPCablecom E-MTA	IPCablecom E-MTA
GSM EFR	Toll quality	Toll quality	Good for speech	Proprietary VQE – Voice quality enhancement	High	Cellular	GSM and 3GPP

Table I.1 – Narrow-band codecs (part 3)

Codec	Performance for background noise	MOS reduction for tandem encodings (Note 13)	Performance	Other functionality	Prevalence	Applications	Status
TDMA IS-641	< Toll quality	< Toll quality	Good for speech without transcoding		Medium	Cellular	TDMA
EVRC IS-127	< Toll quality	< Toll quality	Good for speech without transcoding		High	Cellular	CDMA/3GPP2
IS-733	Toll quality	Toll quality	Good for speech without transcoding		High	Cellular	CDMA/3GPP2
AMR	Toll quality (at 12.2 kbit/s)	Toll quality (at 12.2 kbit/s)	Good for speech (at 12.2 kbit/s)	Proprietary VQE	High	Cellular	GSM and 3GPP
IS-893 SMV (Note 4)	Toll quality built-in noise suppression	Toll quality (mode_0)	Good for speech	Noise suppression	Low	Cellular	3GPP2

NOTE 1 – The algorithmic delay is the absolute minimum delay the algorithm will introduce. It is usually the buffering delay of the algorithm. Here it is the sum of the codec frame size and the look ahead.

NOTE 2 – Total codec delay is defined by ITU-T as two times codec frame size plus codec look ahead.

NOTE 3 – RAM usage is reported in 16-bit words, the most common unit for fixed-point DSP implementations (due to 16-bit word length of many common DSPs). Stated RAM usage numbers include: "state memory RAM usage" of the encoder, the "state memory RAM usage" of the decoder and the worst case "temporary RAM usage" of the encoder and the decoder for the TI TMS320C54x architecture.

NOTE 4 – SMV = selectable mode vocoder. Design goals set forth by 3GPP2/TIA are: mode_0, to be better than EVRC under all conditions; and mode_1, to be as good as EVRC.

NOTE 5 – Complexity is reported as MIPS (millions instructions per second) and stated computational complexity numbers include one encoder and one decoder for the TI TMS320C54x architecture; SMV reports as WMOPS.

NOTE 6 – This MOS CQE score is calculated based on Ie values according to [ITU-T G.107], assuming a network delay of 50 ms. This is not the MOS value obtained from listening tests.

NOTE 7 – This MOS CQE score assumes that a call originates and terminates within the same MSO's IP network but travels from coast to coast (worst-case intra-MSO scenario). The network delay is assumed to be 90 ms, consisting of 40 ms of propagation delay through various network nodes from coast to coast at 10 ms per 1000 km, and 50 ms of other network delays and jitter.

NOTE 8 – This MOS CQE score assumes that a call originates in one MSO's IP network, goes through an IP backbone network from coast to coast, and then terminates in another MSO's IP network (worst-case intra-MSO scenario). The network delay is assumed to be 140 ms, consisting of 40 ms of propagation delay through various network nodes from coast to coast at 10 ms per 1000 km, and 50 ms of other network delays and jitter for each of the two MSO's IP networks.

Table I.1 – Narrow-band codecs (part 3)

NOTE 9 – This MOS CQE score assumes that a call originates in one MSO's IP network, goes through PSTN from coast to coast, and then terminates in a traditional land-line telephone (worst-case cable-to-PSTN scenario). The network delay is assumed to be 65 ms, consisting of 15 ms of propagation delay through PSTN from coast to coast at near speed of light, and 50 ms of network delays and jitter for the MSO's IP network.

NOTE 10 – This MOS CQE score assumes that a call originates in one MSO's IP network, goes through PSTN from coast to coast, and then terminates in a cellular phone (worst-case cable-to-cellular scenario). The network delay is assumed to be 145 ms, consisting of 15 ms of propagation delay through PSTN from coast to coast at near speed of light, 50 ms of network delays and jitter for the MSO's IP network, and 80 ms of delay going through the cellular phone network.

NOTE 11 – Per ITU-T listening MOS procedure stipulated in [b-ITU-T P.800], ITU-T Recs P.830 and P.831, it is required to have known reference standards in the same tests. Scores between two codecs in such a comparison provides a greater relative performance indication.

<<: denotes that MOS is far away (typically 0.4 or more).

< : denotes that MOS is meaningfully worse (typically 0.2).

~ : denotes that MOS is somewhat less, however, the differences are not statistically meaningful (typically within 0.1).

= : denotes that MOS is really comparable and could actually be higher than the reference codec (but not statistically meaningful).

> : denotes that MOS is better than reference codec.

NOTE 12 – This is the random packet loss rate at which point the codec MOS degrades 0.5 from the clear-channel MOS. The higher the number, the more robust the algorithm is to packet loss. Packet loss concealment algorithms are assumed for all codecs.

NOTE 13 – Tandem encodings means two encodings using the same codec with G.711 in between.

Appendix II

Characteristics of wideband codecs

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

The information provided in this appendix has been gathered by several people from a variety of sources. As such, it is possible that errors have been made. It is provided for general information. Please consult the appropriate source documents before using it for developmental purposes.

Table II.1 – Wideband codecs (part 1)

Codec	Technology	Speech model used	Year standardized	Sampling rate	Audio bandwidth	Coding rate (kbit/s)	Source controlled variable coding rate	Codec frame size (ms)	Look ahead (ms)
G.722	SB-ADPCM (sub-band ADPCM – two sub-bands 0-4 kHz and 4-8 kHz)	No	1988	16 kHz	50 Hz to 7 kHz	48, 56, 64	No	0.0625	0
G.722.1	Modulated lapped transform (MLT)	No	1999	16 kHz	50 Hz to 7 kHz	24, 32	No	20	20
G.722.2/ AMR-WB	ACELP	Yes	2001/2002	16 kHz	50 Hz to 6.4 kHz (Note 1)	6.60, 8.85, 12.65, 14.25, 15.85, 18.25, 19.85, 23.05, 23.85	Defined in [ETSI TS 126 093]	20	5
VMR-WB	ACELP	Yes	2004	16 kHz	60 Hz to 6.4 kHz (Note 2)	Rate Set I: 8.55, 4.0, 0.8 Rate Set II: 13.3, 6.2, 2.7, 1.0	Yes	20	11.875
SMV-WB	eX-CELP with an efficient rate determination algorithm	Yes	Not an approved standard	16 kHz	60 Hz to 7 kHz	Three modes: 0: Avg. 9.0 1: Avg. 7.7 2: Avg. 6.2	Yes	20	9.75
iPCM-wb (GIPS)	Multiple description waveform coding	No	Not an approved standard	16 kHz	50 Hz to > 7 kHz (Note 8)	Avg. 80	Yes	10, 20, 30, 40	0
iSAC (GIPS)	Transform coding	No	Not an approved standard	16 kHz	50 Hz to > 7 kHz (Note 8)	Variable 10-32	Yes	30 or 60 (adaptive)	3
BV32 (Broadcom)	TSNFC (two-stage noise feedback coding, same as BV16)	Yes	Not an approved standard	16 kHz	20 Hz to > 7 kHz (Note 8)	32	No	5	0

Table II.1 – Wideband codecs (part 2)

Codec	Encoder and decoder filtering delay (ms) (Note 3)	Algorithmic delay (ms) (Note 4)	Total codec delay (ms) (Note 5)	RAM (kwords)	Total memory footprint (kwords) (Note 6)	Complexity	Output quality versus G.722 for clean speech (Note 7)	Packet loss rate for 0.5 MOS degradation	Performance for music
G.722	1.4375	1.5	1.5625	1	6	10 MIPS	(Self)	No PLC mechanism specified	64 kbit/s: okay for audio 48 kbit/s: marginal quality (noisy)
G.722.1	0	40	60	5.5	15.7	10.3 WMOPS	24 kbit/s < G.722 @ 56 kbit/s; 24 kbit/s ≥ G.722 @ 48 kbit/s; 32 kbit/s < G.722 @ 64 kbit/s; (32 kbit/s ≥ G.722 @ 56 kbit/s except tandeming and –16 dBov)	1.5% for G.722.1 @ 24 kbit/s 1.0% for G.722.1 @ 32 kbit/s	Relatively good for music
G.722.2/AMR-WB	1.875	26.875	46.875	5.3	23	38 WMOPS	6.6, 8.85 kbit/s < G.722 @ 48 15.85 kbit/s ≥ G.722 @ 56 23.85 kbit/s ≥ G.722 @ 64	2.0% for 12.65 and 15.85 kbit/s 1.8% for 19.85 and 23.85 kbit/s	Not good but no annoying effects (Note 10)
VMR-WB	1.875	33.75	53.75	9.05	40	42 WMOPS	Mode 0 = G.722.2 @ 14.25 Mode 1 = G.722.2 @ 12.65 except tandeming Mode 2 > G.722.2 @ 8.85 Mode 3 = G.722 @ 56 Mode 4 > G.722.2 @ 8.85 (Note 9)	2.2% for mode 0 2.2% for mode 1 2.5% for mode 2 (Note 9)	Not good but no annoying effects
SMV-WB	2.5	32.25	52.25	9.81	30.8	38 MIPS	Mode 0 = G.722.2 @ 14.25 except tandeming Mode 1 = G.722.2 @ 12.65 except tandeming Mode 2 > G.722.2 @ 8.85	1.4% for mode 0 1.8% for mode 1 2.0% for mode 2 (Note 9)	Okay for audio with music detection algorithm
iPCM-wb (GIPS)	0	10, 20, 30, 40	20, 40, 60, 80	4.6	19.4	8.6 MIPS	≥ G.722 > iSAC and G.722.2	22.0%	Very good
iSAC (GIPS)	0	33 or 63 (adaptive)	63 or 123 (adaptive)	~ 4.4	~ 30	Claimed equivalent to G.722.2	≥ G.722.2	2-5% depending on rate	Good
BV32 (Broadcom)	0	5	10	3	13	17.5 MIPS	≥ G.722 @ 64 kbit/s	5.3%	Okay for audio

Table II.1 – Wideband codecs (part 3)

Codec	Other functionality	Prevalence in wideband-capable systems *	Integrated packet loss concealment	Integrated VAD/DTX/CNG	Reference fixed point C-code
G.722		High	No	No	Available
G.722.1		Low	Frame repeat specified in G.722.1	No	Available
G.722.2/AMR-WB		Mandatory 3GPP WB codec soon to be deployed	Yes	Yes	Available
VMR-WB	Integrated noise suppression	Recently standardized by 3GPP2	Yes	Yes	Available
SMV-WB	Integrated noise suppression	Low	Yes	Yes	Available
iPCM-wb (GIPS)	Noise cancellation/suppression available with VQE/NetEQ	Low	Optional VQE/NetEQ module	Optional with VQE/NetEQ module	Available
iSAC (GIPS)	Noise cancellation/suppression available with VQE/NetEQ	Very high (Skype, Google, AOL, Yahoo, QQ, etc.)	Optional VQE/NetEQ module	Optional with VQE/NetEQ module	Available
BV32 (Broadcom)		Low	Yes	No	Available

NOTE 1 – G.722.2/AMR-WB only encodes the 0-6.4 kHz band; the 6.4-7 kHz band is estimated and synthesized based on low-band information, except the 23.85 kbit/s mode also encodes 6.4-7 kHz band energy (the other modes do not).

NOTE 2 – VMR-WB only encodes 0-6.4 kHz band; the 6.4-7 kHz band is estimated and synthesized based on low-band information.

NOTE 3 – Encoder and decoder filtering delay may include the delays caused by low-pass filtering in sampling rate conversion and by analysis and synthesis filterbanks in sub-band coding approaches.

NOTE 4 – The algorithmic delay is the absolute minimum delay the algorithm will introduce. It is usually the buffering delay of the algorithm. Here it is the sum of codec frame size, look ahead, and filtering delay.

NOTE 5 – Total codec delay here is defined as two times codec frame size plus look ahead and filtering delay.

NOTE 6 – Total memory footprint is the total memory size required for implementing a single channel of full-duplex codec. This number is the sum of the memory sizes for the program, data tables, and data RAM, where data RAM includes scratch RAM (re-usable work space, dynamic memory) and instance memory (static memory that needs to be carried over from one frame to the next). This is a representative number and can vary slightly depending on the processor.

NOTE 7 – Includes IPR-holder claims not independently verified.

NOTE 8 – The iPCM, ISAC and BV32 codecs claim an upper frequency response close to 8 kHz.

NOTE 9 – The operating modes 0, 1, 2 and 4 are specific to 3GPP2 while mode 3 is interoperable with the AMR-WB codec at 12.65 kbit/s. Modes 0-3 apply to rate set II, while mode 4 is defined for rate set I.

NOTE 10 – HiFi extension available through AMR-WB+.

Bibliography

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

[b-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*.

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

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

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

[b-ITU-T H.264] ITU-T Recommendation H.264 (2005), *Advanced video coding for generic audiovisual services*.

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

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

[b-ITU-T J.360] ITU-T Recommendation J.360 (2006), *IPCablecom2 Architecture Framework*.

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

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

[b-ITU-T P.800] ITU-T Recommendation P.800 (1996), *Methods for subjective determination of transmission quality*.

[b-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*.

[b-ANSI/TIA 825-A] ANSI/TIA-825-A-2003 (2003), *A Frequency Shift Keyed Modem for Use on the Public Switched Telephone Network*.

[b-ARIB 27H] ARIB RCR STD-27H (2005), *Personal Digital Cellular Telecommunication System*.

[b-DESK CONF] *Aldred* (B.K.): Desktop Conferencing: A Complete Guide to Its Applications and Technology. McGraw-Hill, 1996.

[b-ETSI TR 102 493] ETSI TR 102 493 V1.1.1 (2005), *Speech Processing, Transmission and Quality Aspects (STQ); Guidelines for the use of Video Quality Algorithms for Mobile Applications*.

[b-ETSI TS 126 141] ETSI TS 126 141 V7.1.0 (2007), *Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications Systems (UMTS); IP Multimedia System (IMS) Messaging and Presence; Media formats and codecs*.

[b-ETSI TR 126 936] ETSI TR 126 936 V6.1.0 (2006), *Universal Mobile Telecommunications Systems (UMTS); Performance characterization of 3GPP audio codecs*.

[b-3GPP TS 24.229] 3GPP TS 24.229 V7.6.0 (2006), *Internet Protocol (IP) multimedia call control protocol based on Session Initiation Protocol (SIP) and Session Description Protocol (SDP); Stage 3*.

[b-3GPP TS 26.090] 3GPP TS 26.090 V6.0.0 (2005), *AMR Speech Codec; Transcoding Functions*.

[b-3GPP TS 26.911] 3GPP TS 26.911 V7.1.0 (2006), *Codec(s) for Circuit-Switched (CS) multimedia telephony service; Terminal implementor's guide*.

[b-3GPP TS 44.018] 3GPP TS 44.018 V7.6.0 (2006), *Mobile radio interface layer 3 specification; Radio Resource Control (RRC) protocol*.

[b-3GPP TS 46.060] 3GPP TS 46.060 V6.0.0 (2005), *Enhanced Full Rate (EFR) speech transcoding*.

[b-IEEE COM1] IEEE Communications Magazine – *Extended AMR-WB for High-Quality Audio on Mobile Devices*, R. Salami *et al.*, May 2006.

[b-IETF RFC 2190] IETF RFC 2190 (1997), *RTP Payload Format for H.263 Video Streams*.

[b-IETF RFC 3261] IETF RFC 3261 (2002), *SIP: Session Initiation Protocol*.

[b-IETF RFC 3435] IETF RFC 3435 (2003), *Media Gateway Control Protocol (MGCP) Version 1.0*.

[b-IETF RFC 3890] IETF RFC 3890 (2004), *A Transport Independent Bandwidth Modifier for Session Description Protocol (SDP)*.

[b-ISMA-2] Internet Streaming Media Alliance, *ISMA Implementation Specification 2.0*, April 2005.

[b-ISO/IEC 14496-3 Amd.2] ISO/IEC 14496-3:2005/Amd.2:2006, *Coding of audio-visual objects – Part 3: Audio – Amendment 2, Audio Lossless Coding (ALS), new profiles and BSAC extensions*.

[b-ID MGCP] IETF Internet Draft (2005), *RTCP XR VoIP Metrics Package for the Media Gateway Control Protocol*, draft-auerbach-mgcp-rtcpxr-03.

[b-ID MGCP FAX] IETF Internet Draft (2005), *Media Gateway Control Protocol Fax Package*, draft-andreasen-mgcp-fax-04.

[b-MUSHRA] ITU-R Recommendation BS.1534-1 (2003), *Method for the subjective assessment of intermediate quality levels of coding systems*.

[b-Telcordia GR 506] Telcordia GR-506 (2006), *LSSGR: Signalling for Analogue Interfaces*.

[b-TIA/EIA 136-410] TIA/EIA-136-410 (2001), *Enhanced Full Rate Voice Codec*.

[b-ID VOIP] IETF Internet Draft (2005), *VoIP Metrics Management Information Base*, draft-ietf-avt-rtcp-xr-mib-03.txt.

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