

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 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/>.

© ITU 2008

All rights reserved. No part of this publication may be reproduced, by any means whatsoever, without the prior written permission of ITU.

CONTENTS

Page

1	Scope and introduction	1
1.1	Scope	1
1.2	Introduction	1
2	References.....	3
3	Terms and definitions	8
4	Abbreviations and acronyms	9
5	Conventions	10
6	Common criteria for media transport over IP	11
6.1	IP network criteria for codec support	11
6.2	Overall quality targets	13
6.3	Media security	13
7	Gap analysis between 3GPP, 3GPP2 and IPCablecom codec requirements	13
7.1	Audio codecs	13
7.2	Video codecs	14
7.3	Quality metrics	15
8	Codec and media requirements.....	15
8.1	RTP requirements	17
8.2	RTCP requirements	17
8.3	General session description for codecs.....	19
8.4	Narrow-band codec specifications	23
8.5	Wideband codec specification	47
8.6	Super-wideband codec specifications.....	56
8.7	Video codec specification.....	65
8.8	Media quality measurement and monitoring	75
	Annex A – H.263 profiles and levels	82
	Annex B – H.264/AVC profiles and levels.....	84
	Appendix I – Characteristics of narrow-band codecs	86
	Appendix II – Characteristics of wideband codecs.....	91
	Bibliography.....	94

Copyright International Telecommunication Union/ITU Telecommunication Sector
Provided by IHS Markit under license with ITU-T
No reproduction or networking permitted without license from IHS

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

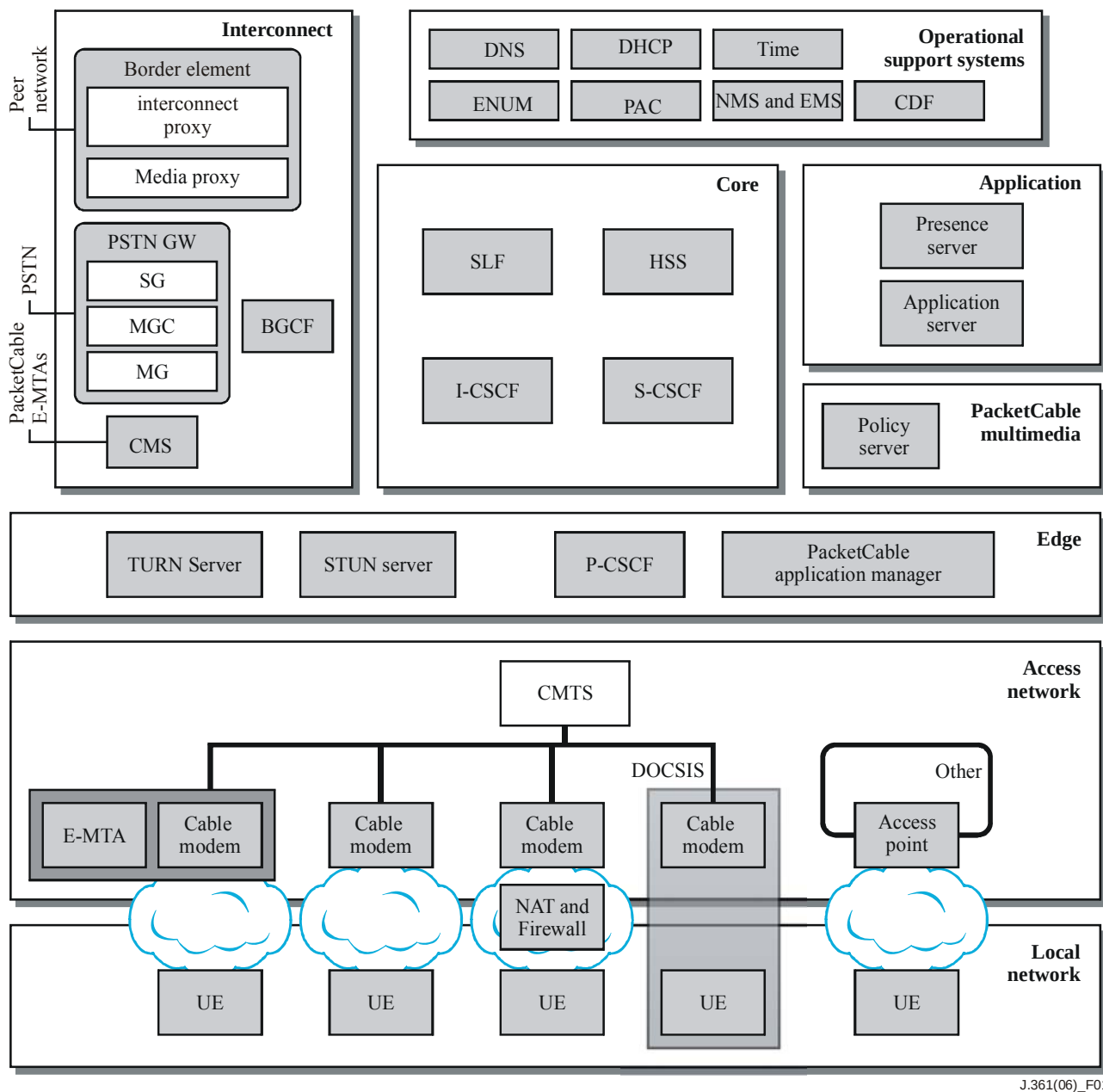
IPCablecom2 defines a modular architecture and a set of interoperable interfaces that leverage emerging communications technologies, such as SIP, to support the rapid introduction of new IP-based communications and streaming services onto the cable network. A modular approach allows operators to deploy 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].



J.361(06)_F01

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), <i>Framework Recommendation for multimedia services</i> .
[ITU-T F.703]	ITU-T Recommendation F.703 (2000), <i>Multimedia conversational services</i> .
[ITU-T F.724]	ITU-T Recommendation F.724 (2005), <i>Service description and requirements for videotelephony services over IP networks</i> .
[ITU-T G.107]	ITU-T Recommendation G.107 (2005), <i>The E-Model, a computational model for use in transmission planning</i> .
[ITU-T G.114]	ITU-T Recommendation G.114 (2003), <i>One-way transmission time</i> .
[ITU-T G.168]	ITU-T Recommendation G.168 (2004), <i>Digital network echo cancellers</i> .
[ITU-T G.711]	ITU-T Recommendation G.711 (1988), <i>Pulse code modulation (PCM) of voice frequencies</i> .
[ITU-T G.711 App.I]	ITU-T Recommendation G.711 Appendix I (1999), <i>A high quality low-complexity algorithm for packet loss concealment with G.711</i> .
[ITU-T G.711 App.II]	ITU-T Recommendation G.711 Appendix II (2000), <i>A comfort noise payload definition for ITU-T G.711 use in packet-based multimedia communication systems</i> .
[ITU-T G.712]	ITU-T Recommendation G.712 (2001), <i>Transmission performance characteristics of pulse code modulation channels</i> .
[ITU-T G.722]	ITU-T Recommendation G.722 (1988), <i>7 kHz audio-coding within 64 kbit/s</i> .
[ITU-T G.722.2]	ITU-T Recommendation G.722.2 (2003), <i>Wideband coding of speech at around 16 kbit/s using Adaptive Multi-Rate Wideband (AMR-WB)</i> .
[ITU-T H.262]	ITU-T Recommendation H.262 (2000) ISO/IEC 13818-2:2000, <i>Information technology – Generic coding of moving pictures and associated information: Video</i> .
[ITU-T T.38]	ITU-T Recommendation T.38 (2005), <i>Procedures for real-time Group 3 facsimile communication over IP networks</i> .
[ITU-T T.140]	ITU-T Recommendation T.140 (1998), <i>Protocol for multimedia application text conversation</i> .
[ITU-T V.18]	ITU-T Recommendation V.18 (2000), <i>Operational and interworking requirements for DCEs operating in the text telephone mode</i> .
[ITU-T V.152]	ITU-T Recommendation V.152 (2005), <i>Procedures for supporting voice-band data over IP networks</i> .
[ANSI/SCTE 24-21]	ANSI/SCTE 24-21 (2006), <i>BV16 Speech Codec Specification for Voice over IP Applications in Cable Telephony</i> .
[ANSI T1.508]	ANSI T1.508 (2003), <i>Loss Plan for Digital Networks</i> .
[ANSI/TIA-127-A]	ANSI/TIA 127-A (2004), <i>Enhanced Variable Rate Codec Speech Service Option 3 for Wideband Spread Spectrum Digital Systems</i> .
[ATSC A/52B]	ATSC A/52B, <i>Digital Audio Compression (AC-3, E-AC-3), Standard, Rev. B</i> , 14 June 2005.

[ETSI TS 102 005]	ETSI TS 102 005 V1.3.1 (2007), <i>Digital Video Broadcasting (DVB); Specification for the use of Video and Audio Coding in DVB services delivered directly over IP protocols.</i>
[ETSI TS 126 091]	ETSI TS 126 091 V6.0.0 (2004), <i>AMR speech codec; Error concealment of lost frames.</i>
[ETSI TS 126 092]	ETSI TS 126 092 V6.0.0 (2004), <i>AMR speech codec; Comfort noise for AMR speech traffic channels.</i>
[ETSI TS 126 093]	ETSI TS 126 093 V6.1.0 (2006), <i>AMR speech codec; Source controlled rate operation.</i>
[ETSI TS 126 094]	ETSI TS 126 094 V6.1.0 (2006), <i>AMR speech codec; Voice activity detector for AMR speech traffic channels.</i>
[ETSI TS 126 101]	ETSI TS 126 101 V6.0.0 (2005), <i>Adaptive Multi-Rate (AMR) speech codec frame structure.</i>
[ETSI TS 126 140]	ETSI TS 126 140 V7.1.0 (2007), <i>Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications Systems (UMTS); Multimedia Messaging Service (MMS); Media formats and codes.</i>
[ETSI TS 126 190]	ETSI TS 126 190 V6.1.1 (2005), <i>Adaptive Multi-Rate – Wideband (AMR-WB) speech codec; Transcoding functions.</i>
[ETSI TS 126 191]	ETSI TS 126 191 V6.0.0 (2004), <i>AMR speech codec, wideband; Error concealment of lost frames.</i>
[ETSI TS 126 192]	ETSI TS 126 192 V6.0.0 (2004), <i>Mandatory speech codec speech processing functions AMR Wideband speech codec; Comfort noise aspects.</i>
[ETSI TS 126 193]	ETSI TS 126 193 V6.1.0 (2006), <i>Adaptive Multi-Rate – Wideband (AMR-WB) speech codec; Source controlled rate operation.</i>
[ETSI TS 126 194]	ETSI TS 126 194 V6.0.0 (2004), <i>Mandatory speech codec speech processing functions AMR Wideband (AMR-WB) speech codec; Voice activity detector (VAD).</i>
[ETSI TS 126 201]	ETSI TS 126 201 V6.0.0 (2004), <i>AMR speech codec, wideband; Frame structure.</i>
[ETSI TS 126 234]	ETSI TS 126 234 V.6.10.0 (2006), <i>Transparent end-to-end packet-switched streaming service (PSS); Protocols and codecs.</i>
[ETSI TS 126 235]	ETSI TS 126 235 V6.4.0 (2005), <i>Packet switched conversational multimedia applications; Default codecs.</i>
[ETSI TS 126 236]	ETSI TS 126 236 V6.4.0 (2005), <i>Packet switched conversational multimedia applications; Transport protocols.</i>
[ETSI TS 126 290]	ETSI TS 126 290 V6.3.0 (2005), <i>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.</i>
[ETSI TS 126 346]	ETSI TS 126 346 V7.6.0 (2008), <i>Universal Mobile Telecommunications Systems (UMTS); Multimedia Broadcast/Multicast Service (MBMS); Protocols and codecs.</i>

- [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), <i>Session Description Protocol (SDP) Simple Capability Declaration</i> .
[IETF RFC 3550]	IETF RFC 3550 (2003), <i>RTP: A Transport Protocol for Real-Time Applications</i> .
[IETF RFC 3551]	IETF RFC 3551 (2003), <i>RTP Profile for Audio and Video Conferences with Minimal Control</i> .
[IETF RFC 3555]	IETF RFC 3555 (2003), <i>MIME Type Registration of RTP Payload Formats</i> .
[IETF RFC 3558]	IETF RFC 3558 (2003), <i>RTP Payload Format for Enhanced Variable Rate Codecs (EVRC) and Selectable Mode Vocoders (SMV)</i> .
[IETF RFC 3611]	IETF RFC 3611 (2003), <i>RTP Control Protocol Extended Reports (RTCP XR)</i> .
[IETF RFC 3640]	IETF RFC 3640 (2003), <i>RTP Payload Format for Transport of MPEG-4 Elementary Streams</i> .
[IETF RFC 3951]	IETF RFC 3951 (2004), <i>Internet Low Bit Rate Codec (iLBC)</i> .
[IETF RFC 3952]	IETF RFC 3952 (2004), <i>Real-time Transport Protocol (RTP) Payload Format for internet Low Bit Rate Codec (iLBC) Speech</i> .
[IETF RFC 3984]	IETF RFC 3984 (2005), <i>RTP Payload Format for H.264 Video</i> .
[IETF RFC 4103]	IETF RFC 4103 (2005), <i>RTP Payload for Text Conversation</i> .
[IETF RFC 4184]	IETF RFC 4184 (2005), <i>RTP Payload Format for AC-3 Audio</i> .
[IETF RFC 4298]	IETF RFC 4298 (2005), <i>RTP Payload Format for BroadVoice Speech Codecs</i> .
[IETF RFC 4348]	IETF RFC 4348 (2006), <i>Real-Time Transport Protocol (RTP) Payload Format for the Variable-Rate Multimode Wideband (VMR-WB) Audio Codec</i> .
[IETF RFC 4352]	IETF RFC 4352 (2006), <i>RTP Payload Format for the Extended Adaptive Multi-Rate Wideband (AMR-WB+) Audio Codec</i> .
[IETF RFC 4424]	IETF RFC 4424 (2006), <i>Real-Time Transport Protocol (RTP) Payload Format for the Variable-Rate Multimode Wideband (VMR-WB) Extension Audio Codec</i> .
[IETF RFC 4566]	IETF RFC 4566 (2006), <i>SDP: Session Description Protocol</i> .
[IETF RFC 4598]	IETF RFC 4598 (2006), <i>Real-time Transport Protocol (RTP) Payload Format for Enhanced AC-3 (E-AC-3) Audio</i> .
[ISO/IEC 11172-3]	ISO/IEC 11172-3:1993, <i>Information technology – Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s – Part 3: Audio</i> .
[ISO/IEC 13818-3]	ISO/IEC 13818-3:2007, <i>Information technology – Generic coding of moving pictures and associated audio information – Part 3: Audio</i> .
[ISO/IEC 13818-7]	ISO/IEC 13818-7:2006, <i>Information technology – Generic coding of moving pictures and associated audio information – Part 7: Advanced Audio Coding (AAC). (MPEG2)</i>
[ISO/IEC 14496-1]	ISO/IEC 14496-1:2004, <i>Information technology – Coding of audio-visual objects – Part 1: Systems. (MPEG4)</i>

[ISO/IEC 14496-2]	ISO/IEC 14496-2: 2004, <i>MPEG-4, Coding of audio-visual objects – Part 2: Visual</i> .
[ISO/IEC 14496-3]	ISO/IEC 14496-3:2005, <i>Information technology – Coding of audio-visual objects – Part 3: Audio. (MPEG-4)</i>
[TIA/IS-893]	TIA/IS-893 (2005), <i>Selectable Mode Vocoder (SMV) Service Option for Wideband Spread Spectrum Communication Systems</i> .
[TIA-1016-A]	TIA-1016-A (2006), <i>Source-Controlled Variable-Rate Multimode Wideband Speech Codec (VMR-WB), Service Options 62 and 63 for Spread Spectrum Systems</i> .

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 IP-Cablecom 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 IP-Cablecom 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 IP-Cablecom 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 IP-Cablecom 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 IPCom2 device to support every possible codec technology that could be interconnected with an IPCom system. Transcoding within the IPCom 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 [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 IP-Cablecom

One of the mandates of IP-Cablecom2 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, IP-Cablecom2 codec selection includes additional codecs beyond those specified by 3GPP and 3GPP2. IP-Cablecom 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 IP-Cablecom 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 IP_Cablecom2

IP_Cablecom2 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. IP_Cablecom2 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 IP_Cablecom

IP_Cablecom 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. IP_Cablecom 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 IP_Cablecom networks and peer IMS networks, including 3GPP or 3GPP2 cellular networks. Additionally, applications based on IP_Cablecom architecture may utilize VoIP metrics to resolve voice quality issues. Therefore, the support of the VoIP metrics block for conversational applications is defined for IP_Cablecom 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 IP_Cablecom 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 IP_Cablecom 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 IPCom network and provides TDM trunk interfaces to the PSTN.

VoIP and video sessions between client devices in the IPCom network and another VoIP network must traverse a media proxy, which acts as a gateway to protect the IPCom 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 IPCom network and a peer VoIP network.

VoIP or video sessions from client devices may also terminate at a media server within the IPCom 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.

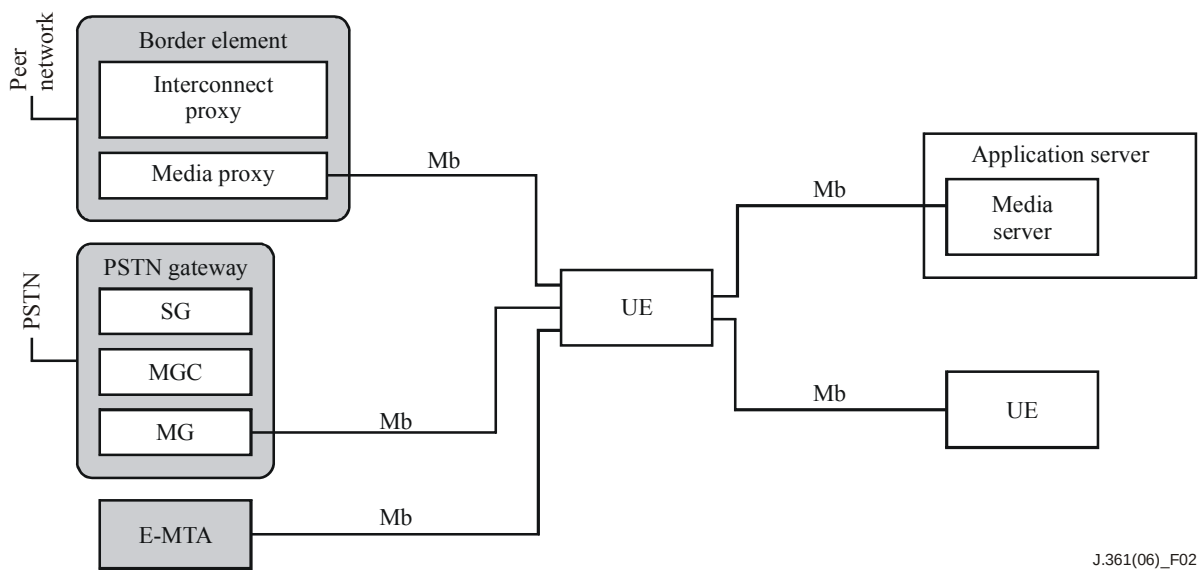


Figure 8-1 – RTP/RTCP media connection

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

Table 8-1 – Media stream interfaces

Reference point	IPCom 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-IP-Cablecom 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 liveness 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 (SDS): 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. IP-Cablecom 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=seqn: <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.

fntp:

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 "fntp" 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=ptime: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=ptime: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=ptime: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=ptime: 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]. IP-Cablecom 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 IP-Cablecom 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