PacketCable™ 2.0

Codec and Media Specification

PKT-SP-CODEC-MEDIA-I03-070925

ISSUED

Notice

This PacketCable specification is the result of a cooperative effort undertaken at the direction of Cable Television Laboratories, Inc. for the benefit of the cable industry and its customers. This document may contain references to other documents not owned or controlled by CableLabs. Use and understanding of this document may require access to such other documents. Designing, manufacturing, distributing, using, selling, or servicing products, or providing services, based on this document may require intellectual property licenses from third parties for technology referenced in this document.

Neither CableLabs nor any member company is responsible to any party for any liability of any nature whatsoever resulting from or arising out of use or reliance upon this document, or any document referenced herein. This document is furnished on an "AS IS" basis and neither CableLabs nor its members provides any representation or warranty, express or implied, regarding the accuracy, completeness, noninfringement, or fitness for a particular purpose of this document, or any document referenced herein.

Document Status Sheet

Document Control Number:	PKT-SP-CODEC-MEDIA-I03-070925			
Document Title:	Codec and Media Specification			
Revision History:	I01 - Released 04/05/06 I02 - Released 10/13/06 I03 - Released 09/25/07			
Date:	September 25, 2007			
Status:	~~Work in Progress~~	~~Draft~~	Issued	~~Closed~~
Distribution Restrictions:	~~Authors Only~~	~~CL/Member~~	~~CL/ Member/ Vendor~~	Public

Key to Document Status Codes:

Work in Progress	An incomplete document, designed to guide discussion and generate feedback, that may include several alternative requirements for consideration.
Draft	A document in specification format considered largely complete, but lacking review by Members and vendors. Drafts are susceptible to substantial change during the review process.
Issued	A stable document, which has undergone rigorous member and vendor review and is suitable for product design and development, cross-vendor interoperability, and for certification testing.
Closed	A static document, reviewed, tested, validated, and closed to further engineering change requests to the specification through CableLabs.

Contents

Figures

Tables

This page left blank intentionally.

1 INTRODUCTION AND SCOPE

This document addresses interfaces for audio and video communication between User Equipment (UE), Media Gateways (MG), and other network elements that process media such as media servers. 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.

1.1 Background

PacketCable 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 TV, and call treatment from the TV
- Mobility services and Integration with Cellular and Wireless Networks - Capabilities such as call handoff and roaming between PacketCable VoIP over wireless LAN and cellular networks
- Multimedia Applications - Capabilities such as QoS-enabled audio and video streaming

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 specification development projects at CableLabs, PacketCable leverages existing open standards wherever possible.

PacketCable is based on the IP Multimedia Subsystem (IMS) as defined by the 3rd Generation partnership Project (3GPP). 3GPP is a collaboration agreement between various standards bodies. The scope of 3GPP is to produce Technical Specifications and Technical Reports for GSM and 3rd Generation (3G) Mobile System networks.

Figure 1 below illustrates the functional components that are included in the PacketCable architecture. Refer to the PacketCable Architecture Framework Technical Report for more detail [ARCH-FRM TR].

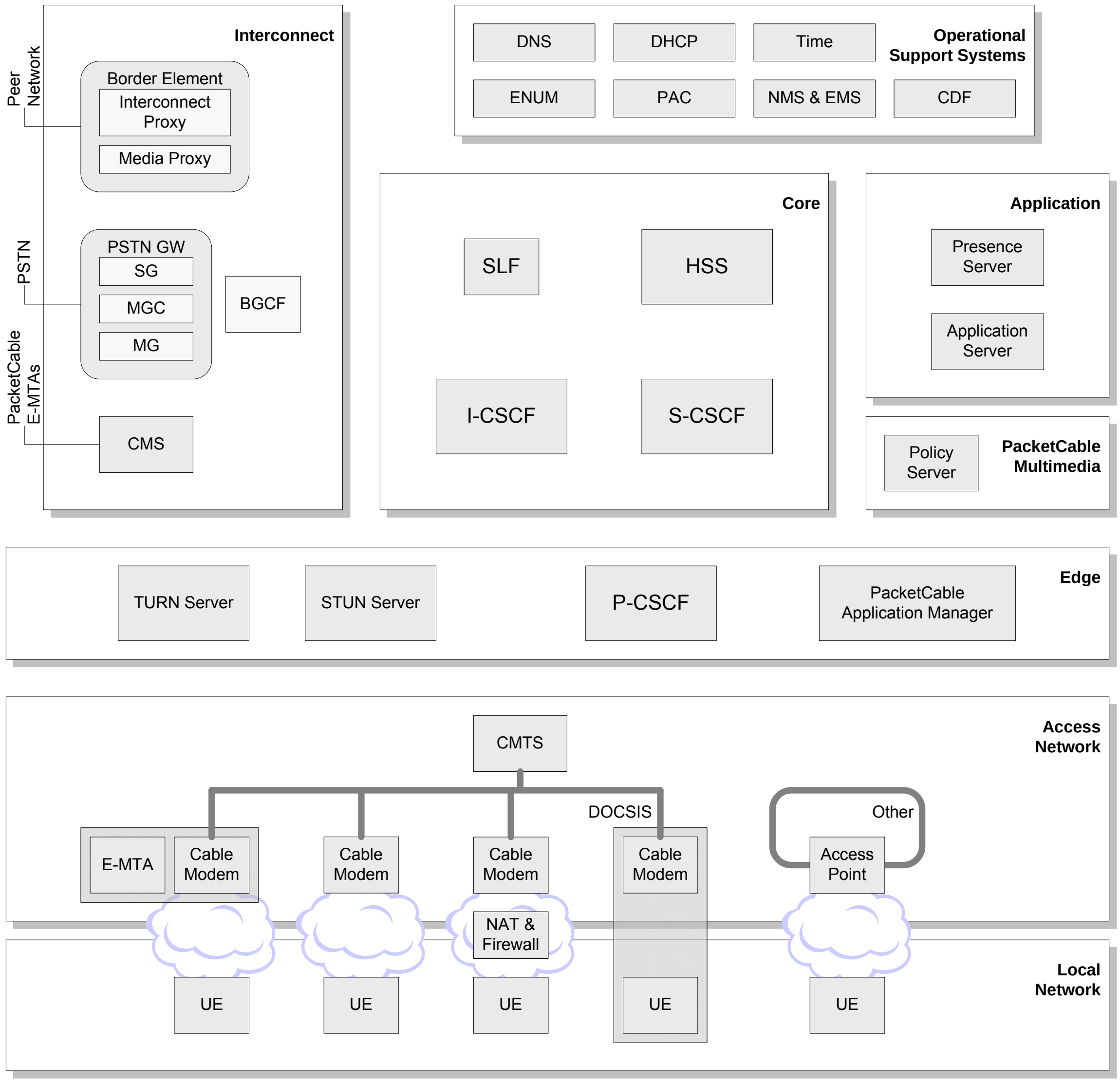

Figure 1 - PacketCable Reference Architecture

Concerning media support, the subject of this specification, the quality of audio and video delivered over the PacketCable 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 document defines codecs and capabilities for supporting narrowband and wideband audio and video applications, with emphasis on the stringent requirements of IP-based voice and video communications.

Acceptable two-way voice and video communications 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 PacketCable 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 PacketCable architecture is to define a robust system that accommodates evolving technology without creating a legacy burden.

1.2 Purpose of the Document

The purpose of this document is to specify profiles for codecs, packetization rules, encodings, and quality metrics to assure successful media interworking within a PacketCable network and between a PacketCable network and interconnecting networks including the PSTN and cellular networks. The media specified in this document includes narrowband audio, wideband audio, image, and video for communications services.

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

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

1.3 High-Level Requirements for PacketCable

PacketCable media stream transport and encoding design goals include:

- Minimize the effects of latency, packet loss and jitter on sensitive media streams (e.g., voice and video) to ensure a quality level 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 wide-band voice codec technologies;
- Accommodate emerging video codec technologies to provide support for applications like video telephony, IP video streaming, etc.;
- 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 TTY;
- Support calculation and reporting of voice quality metrics.

1.4 Organization of Document

This document is organized as follows:

- Scope and High-Level Requirements (Section 1);
- References (Section 2);
- Terms and Definitions (Section 3);
- Abbreviations and Acronyms (Section 4);
- Common Criteria for Transport of Audio and Video over IP (Section 5);
- Gap Analysis between 3GPP, 3GPP2 and PacketCable Requirements (Section 6);
- Codec and Media Requirements (Section 7);
- Video Codec Profiles and Levels (Annex A and Annex B).

1.5 Requirements Syntax

Throughout this document, 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 specification.
"MUST NOT"	This phrase means that the item is an absolute prohibition of this specification.

"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 behavior is acceptable or even useful, but the full implications should be understood and the case carefully weighed before implementing any behavior described with this label.
"MAY"	This word means that this item is truly optional. One vendor may choose to include the item because a particular marketplace requires it or because it enhances the product, for example; another vendor may omit the same item.

2 REFERENCES

2.1 Normative References

In order to claim compliance with this specification, it is necessary to conform to the following standards and other works as indicated, in addition to the other requirements of this specification. Intellectual property rights may be required to use or implement such normative references.

[ATSC A/52B]	ATSC A/52B, Digital Audio Compression Standard (AC-3, E-AC-3), Revision B, June 2005.
[BVOICE]	BroadVoice™16 Speech Codec Specification, Version 1.2, Chen, October 2003.
[G.107]	ITU-T Rec. G.107, The E Model, a computational model for use in transmission planning, March 2005.
[G.114]	ITU-T Rec. G.114, One-way transmission time, May 2003.
[G.168]	ITU-T Rec. G.168, Digital Network Echo Cancellers, August 2004.
[G.711]	ITU-T Rec. G.711, Pulse Code Modulation (PCM) of Voice Frequencies, November 1988.
[G.711-I]	ITU-T Rec. G.711 Appendix I, A high quality low complexity algorithm for packet loss concealment with G.711, September 1999.
[G.711-II]	ITU-T Rec. G.711 Appendix II, A comfort noise payload definition for ITU-T G.711 use in packet-based multimedia communication systems, February 2000.
[G.712]	ITU-T Rec. G.712, Transmission Performance Characteristics of Pulse Code Modulation Channels, November 2001.
[G.722.2]	ITU-T Rec. G.722.2, Wideband coding of speech at around 16 kbit/s using Adaptive Multi-Rate Wideband (AMR-WB), July 2003.
[G.722]	ITU-T Rec. G.722.7, kHz audio-coding within 64 kbit/s, November 1988.
[ID SIP RTCP]	IETF Internet-Draft, Session Initiation Protocol Package for Voice Quality Reporting Event, draft-ietf-sipping-rtcp-summary-02, May 2007, work in progress.
[iLBC]	iLBC Speech Codec Fixed Point Reference Code for PacketCable, Version 1.0.3, October 2003.
[ISO 11172-3]	ISO-IEC 11172-3, Information Technology - Generic coding of moving pictures and associated audio information - Part 3: Audio, August 1993.
[ISO 13818-2]	ISO/IEC 13818-2 MPEG-2, Generic Coding of Moving Pictures and Associated Audio Information—Part 2, November 1994.
[ISO 13818-3]	ISO-IEC 13818-3, Information Technology - Coding of moving pictures and associated audio information - Part 3: Audio, April 1998.
[ISO 13818-7]	ISO-IEC 13818-7 MPEG-2, Generic Coding of Moving Pictures and Associated Audio Information - Part 7: Advanced Audio Coding (AAC), January 2006.
[ISO 14496-1]	ISO-IEC 14496-1 MPEG-4, Coding of Audio-Visual Objects - Part 1: Systems, November 2004.
[ISO 14496-2]	ISO-IEC 14496-2 MPEG-4, Coding of Audio-Visual Objects—Part 2: Visual, January 1999.
[ISO 14496-3]	ISO-IEC 14496-3 MPEG-4, Coding of Audio-Visual Objects - Part 3: Audio, December 2005.
[RFC 3264]	IETF RFC 3264, An Offer/Answer with Session Description Protocol (SDP), June 2002.
[RFC 4629]	IETF RFC 4629, RTP Payload Format for ITU-T Rec. H.263 Video, November 2006.
[RFC 2198]	IETF RFC 2198, RTP Payload for Redundant Audio Data, September 1997.
[RFC 2250]	IETF RFC 2250, RTP Payload Format for MPEG1/MPEG2 Video, January 1998.

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

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

[RFC 3267] IETF RFC 3267, 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, June 2002.

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

[RFC 3550] IETF RFC 3550/STD0064, RTP: A Transport Protocol for Real-Time Applications, July 2003.

[RFC 3551] IETF RFC 3551/STD0065, RTP Profile for Audio and Video Conferences with Minimal Control, July 2003.

[RFC 3555] IETF RFC 3555, MIME Type Registration of RTP Payload Formats, July 2003.

[RFC 3556] IETF RFC 3556, Session Description Protocol (SDP) Bandwidth Modifiers for RTP Control Protocol (RTCP) Bandwidth, July 2003.

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

[RFC 3611] IETF RFC 3611, Real Time Control Protocol Extended Reports (RTCP XR), November 2003.

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

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

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

[RFC 3984] IETF RFC 3984, RTP Payload Format for H.264 Video, February 2005.

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

[RFC 4298] IETF RFC 4298, RTP Payload Format for BroadVoice Speech Codecs, December 2005.

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

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

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

[RFC 4566] IETF RFC 4566 SDP: Session Description Protocol, July 2006.

[RFC 4598] IETF RFC 4598, Real-time Transport Protocol (RTP) Payload Format for Enhanced AC-3 (E-AC-3) Audio, July 2005.

[RFC 4733] IETF RFC 4733, RTP Payload for DTMF Digits, Telephony Tones, and Telephony Signals, December 2006.

[RFC 4734] IETF RFC 4734, Definition of Events for Modem, Fax, and Text Telephony Signals, December 2006.

[T.38] ITU-T Rec. T.38, Procedures for Real-Time Group 3 Facsimile Communication over IP Networks, April 2004.

[T1.508] ANSI T1.508-2003, Loss Plan for Digital Network, December 2003.

[TS 26.090] 3GPP TS 26.090 v7.0.0, Adaptive Multi-Rate (AMR) Speech Codec; Transcoding Functions, June 2007.

[TS 26.091] 3GPP TS 26.091 v7.0.0, Adaptive Multi-Rate (AMR) Speech Codec; Error Concealment of Lost Frames, June 2007.

[TS 26.092] 3GPP TS 26.092 v7.0.0, Adaptive Multi-Rate (AMR) Speech Codec; Comfort Noise Aspects, June 2007.

[TS 26.093] 3GPP TS 26.093 v7.1.0, Adaptive Multi-Rate (AMR) Speech Codec; Source Controlled Rate Operation, June 2007

[TS 26.094] 3GPP TS 26.094 v7.0.0, Adaptive Multi-Rate (AMR) Speech Codec; Voice Activity Detector (VAD), June 2007.

[TS 26.101] 3GPP TS 26.101 v7.0.0, Adaptive Multi-Rate (AMR) Speech Codec Frame Structure, June 2007.

[TS 26.140] 3GPP TS 26.140 v7.1.0, Multimedia Messaging Service (MMS); Media formats and codecs, June 2007.

[TS 26.190] 3GPP TS 26.190 v7.0.1, Adaptive Multi-Rate - Wideband (AMR-WB) Speech Codec; Transcoding Functions, June 2007.

[TS 26.191] 3GPP TS 26.191 v7.0.0, Adaptive Multi-Rate - Wideband (AMR-WB) Speech Codec; Error Concealment of Erroneous or Lost Frames, June 2007.

[TS 26.192] 3GPP TS 26.192 v7.0.0, Adaptive Multi-Rate - Wideband (AMR-WB) Speech Codec; Comfort Noise Aspects, June 2007.

[TS 26.193] 3GPP TS 26.193 v7.0.0, Adaptive Multi-Rate - Wideband (AMR-WB) Speech Codec; Source Controlled Rate Operation, June 2007.

[TS 26.194] 3GPP TS 26.194 v7.0.0, Adaptive Multi-Rate - Wideband (AMR-WB) Speech Codec; Voice Activity Detector (VAD), June 2007.

[TS 26.201] 3GPP TS 26.201 v7.0.0, Adaptive Multi-Rate - Wideband (AMR-WB) Speech Codec; Frame Structure, June 2007.

[TS 26.234] 3GPP TS 26.234 v.7.3.1, Transparent End-to-End Packet-Switched Streaming Service—Protocols and Codecs, July 2007.

[TS 26.235] 3GPP TS 26.235 v7.1.0, Packet Switched Conversational Multimedia Applications; Default Codecs (Release 6), March 2006.

[TS 26.236] 3GPP TS 26.236 v7.1.0 (2005-06), Packet switched conversational multimedia applications; Transport protocols, October 2006.

[TS 26.290] 3GPP TS 26.290 v.7.0, Extended Adaptive Multi-Rate - Wideband (AMR-WB+) Codec; Transcoding Functions, March 2007.

[TS 26.346] 3GPP TS 26.346 v7.4.0, Multimedia Broadcast/Multicast Service (MBMS); Protocols and codecs, June 2007.

[TS 26.401] AACPlus3GPP TS 26.401 v7.0.0, Enhanced AACPlus General Audio Codec, General Description, June 2006.

[TS 26.402] AACPlus3GPP TS 26.402 v7.0.0, Enhanced AACPlus General Audio Codec, Additional Decoder Tools, June 2007.

[TS 26.403] AACPlus 3GPP TS 26.403 v7.0.0, Enhanced AACPlus General Audio Codec, Encoder Specification AAC part, June 2006.

[TS 26.404] AACPlus3GPP TS 26.404 v.7.0.0, Enhanced AACPlus General Audio Codec, Encoder Specification SBR part, June 2007.

[TS 26.405] AACPlus3GPP TS 26.405 v7.0.0, Enhanced AACPlus General Audio Codec, Encoder Specification Parametric Stereo part, June 2007.

[TS 26.410] AACPlus3GPP TS 26.410 v7.1.0, Enhanced AACPlus General Audio Codec, Floating-point ANSI C-code, December 2006.

[TS 26.411] AACPlus3GPP TS 26.411 v7.2.0, Enhanced AACPlus General Audio Codec, Fixed-point ANSI C-code, March 2007.

[TS 102 005]	ETSI TS 102 005 v1.2.1, Digital Video Broadcast (DVB); Specification for the use of Video and Audio Coding in DVB services delivered directly over IP protocols, May 2007.
[TS C.S0014]	3GPP2 C.S0014-0 v1.0, Enhanced Variable Rate Codec, Speech Service Option 3 for Wideband Spread Spectrum Digital Systems, January 1997.
[TS C.S0030]	3GPP2 C.S0030-0 v3.0, Selectable Mode Vocoder (SMV) Service Option for Wideband Spread Spectrum Communication Systems, January 2004.
[TS C.S0052-A]	3GPP2 C.S0052-A version 1.0, Source-controlled variable-rate multimode wideband speech codec (VMR-WB), service options 62 and 63 for spread spectrum systems, April 2005.
[V.18]	ITU-T Rec. V.18, Operational and Interworking Requirements for DCEs Operating in the Text Telephone Mode, November 2000.
[V.152]	ITU-T Rec. V.152, Procedures for supporting Voice-Band Data over IP Networks, January 2005.

2.2 Informative References

This specification uses the following informative references.

[ANSI 136-C]	ANSI/TIA/EIA-136-Rev.C, part 410 - TDMA Cellular/PCS - Radio Interface, Enhanced Full Rate Voice Codec (ACELP) (formerly IS-641), June 2001.
[ANSI 825-A]	ANSI TIA-825-A-2003, A Frequency Shift Keyed Modem for Use on the Public Switched Telephone Network, April, 2003.
[ARCH-FRM TR]	PacketCable Architecture Framework Technical Report, PKT-TR-ARCH-FRM-V03-070925, September 25, 2007, Cable Television Laboratories, Inc.
[ATSC]	ATSC IS/191, ATSC Implementation Subcommittee Finding: Relative timing of sound and vision for broadcast operations, June 2003.
[ARIB 27H]	ARIB, RCR STD-27H - Personal Digital Cellular Telecommunication System RCR Standard, Association of Radio Industries and Businesses (ARIB).
[ETSI TR 102 493]	ETSI TR 102 493 v1.1.1, Speech Processing, Transmission and Quality Aspects (STQ): Guidelines for the Use of Video Quality Algorithms for Mobile Applications, August 2005.
[G.100]	ITU T Rec. G.100/P.10, Vocabulary for performance and Quality of Service, July 2006.
[G.109]	ITU-T Rec. G.109, Definition of categories of speech transmission quality, September 1999.
[G.121]	ITU-T Rec. G.121, Loudness ratings (LRs) of national systems, March 1993.
[G.131]	ITU-T Rec. G.131, Talker echo and its control, November 2003.
[G.729-A]	ITU-T Rec. G.729 Annex A, Reduced Complexity 8 kbit/s CS-ACELP Speech Codec, November 1996.
[G.729-E]	ITU-T Rec. G.729 Annex E, 11.8 kbit/s CS-ACELP speech coding algorithm, September 1998.
[GR 506]	Telcordia GR-506-CORE, Issue 1, Revision 1, LSSGR: Signaling for Analog Interfaces, November 1996.
[G.1050]	ITU-T Rec. G.1050, Network model for evaluating multimedia transmission performance over Internet Protocol, November 2005.
[H.243]	ITU-T Rec. H.243, Procedures for establishing communication between three or more audiovisual terminals using digital channels up to 1920 kbit/s, February 2000.
[H.245]	ITU-T Rec. H.245, Control protocol for multimedia communication, July 2003.
[H.261]	ITU-T Rec. H.261, Video codec for audiovisual services at p x 64 kbit/s, March 1993.
[H.263 Annex X]	ITU-T Rec. H.263—Annex X, Profiles and Levels Definition, January 2005.
[H.263]	ITU-T Rec. H.263, Video coding for low bit rate communication, January 2005.

[H.264]	ITU-T Rec. H.264, Advanced Video Coding for Generic Audiovisual Services, March 2005.
[H.323]	ITU-T Rec. H.323, Packet-based multimedia communications systems, July 2003.
[H.324]	ITU-T Rec. H.324, Terminal for low bit-rate multimedia communication, March 2002.
[ID MGCP FAX]	IETF Internet-Draft Media Gateway Control Protocol Fax Package, draft-andreason-mgcp-fax-06, June, 2007, work in progress.
[ID MGCP]	IETF Internet-Draft RTCP XR VoIP Metrics Package for the Media Gateway Control Protocol, draft-auerbach-mgcp-rtcpxr-06, April 2006, work in progress.
[ID VOIP]	IETF Internet-Draft VoIP Metrics Management Information Base, draft-ietf-avt-rtcp-xr-mib-06.txt, March 2007, work in progress.
[IEEE COM1]	IEEE Communications Magazine, Extended AMR-WB for High-Quality Audio on Mobile Devices, R. Salami et. al., May 2006.
[ISMA-2]	Internet Streaming Media Alliance, ISMA Implementation Specification 2.0, April 2005.
[ISO 14496-3-Amd2]	ISO-IEC 14496-3 MPEG-4, Coding of Audio-Visual Objects - Part 3: Audio, Amendment 2, Audio Lossless Coding (ALS), new profiles and BSAC extensions, March 2006.
[ITUR BS 1534-1]	ITU-R BS. 1534-1, Method for the subjective assessment of intermediate quality level of coding systems, 2003.
[P.56]	ITU-T Rec. P.56, Objective Measurement of Active Speech Level, 1993.
[P.561]	ITU-T Rec. P.561, In-Service, Non-Intrusive Measurement Device, 1996.
[P.800]	ITU-T Rec. P.800, Methods for subjective determination of transmission quality, August 1996.
[P.862]	ITU-T Rec. P.862, Perceptual Estimation of Speech Quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs, February 2001.
[RFC 2190]	IETF RFC 2190, RTP Payload Format for H.263 Video Streams, September 1997.
[RFC 3261]	IETF RFC 3261, SIP: Session Initiation Protocol, June 2002.
[RFC 3435]	IETF RFC 3435, Media Gateway Control Protocol (MGCP) Version 1.0, January 2003.
[RFC 3890]	IETF RFC 3890, A Transport Independent Bandwidth Modifier for Session Description Protocol, September 2004.
[SEC TR]	PacketCable Security Technical Report, PKT-TR-SEC-V03-070925, September 25, 2007, Cable Television Laboratories, Inc.
[TS 24.229]	3GPP TS 24.229 v7.8.0, IP multimedia call control protocol based on Session Initiation Protocol (SIP) and Session Description Protocol (SDP); Stage 3, June 2007.
[TS 26.141]	3GPP TS 26.141 v7.1.0, IP Multimedia System (IMS) Messaging and Presence; Media formats and codecs, March 2006.
[TS 26.936]	3GPP TS 26.936 v7.0.0, Performance characterization of 3GPP audio codecs, June 2007.
[TS 44.018]	3GPP TS 44.018 v7.9.0 - Radio Resource Control (RRC) Protocol, June 2007.
[TS 46.060]	3GPP TS 46.060 v7.0.0, Enhanced Full Rate (EFR) Speech Transcoding, June 2007.
[Y.1540]	ITU-T Rec. Y.1540, Internet protocol data communication service -- IP packet transfer and availability performance parameters, December 2002.
[Y.1541]	ITU-T Rec. Y.1541, Network performance objectives for IP-based services, February 2006.

2.3 Reference Acquisition

- 3GPP specifications: www.3gpp.org
- 3GPP2 specifications: www.3gpp2.org
- ATIS Document center: www.atis.org

- ETSI Standards: www.etsi.org/services_products/e-shop/home.htm
- Internet Engineering Task Force (IETF), Internet: http://www.ietf.org/
  Note: Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time.
  The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt.
- ITU-T Recommendations: www.itu.int/ITU-T/publications/recs.html
- PacketCable specifications: Cable Television Laboratories, www.packetcable.com/specifications/
- Security Industry Association: www.siaonline.org
- Telcordia: http://www.telcordia.com/services/testing/lab_access/gr-listing.html

3 TERMS AND DEFINITIONS

This specification uses the following terms:

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.
Authentication	The act of giving access to a service or device if one has permission to have the access.
Downstream	The direction from the headend toward the subscriber location.
Dynamic Quality of Service	A Quality of Service assigned on the fly for each communication depending on the QoS requested.
Encryption	A method used to translate plaintext into ciphertext.
Endpoint	A Terminal, Gateway or Multipoint Conference Unit (MCU).
Gateway	Devices bridging between the PacketCable 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 Signaling Gateway, which sends and receives circuit switched network signaling to the edge of the PacketCable network.
H.323	An ITU-T recommendation for transmitting and controlling audio and video information. The H.323 recommendation requires the use of the ITU-T H.225 and ITU-T H.245 protocol for communication control between a "gateway" audio/video endpoint and a "gatekeeper" function.
Header	Protocol control information located at the beginning of a protocol data unit.
Internet Engineering Task Force	A body responsible, among other things, for developing standards used on the Internet.
Jitter	Variability in the delay of a stream of incoming packets making up a flow such as a voice communication.
Key	A mathematical value input into the selected cryptographic algorithm.
Latency	The time taken for a signal to pass through a device or network.
Media Gateway	Provides the bearer circuit interfaces to the PSTN and transcodes the media stream.
Network Management	The functions related to the management of data across the network.
Off-Net Call	A communication connecting a PacketCable subscriber out to a user on the PSTN.
One-way Hash	A hash function that has an insignificant number of collisions upon output.
On-Net Call	A communication placed by one customer to another customer entirely on the PacketCable Network.

Privacy	A way to ensure that information is not disclosed to any one other then the intended parties. Information is usually encrypted to provide confidentiality. Also known as confidentiality.
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.
Pulse Code Modulation	A common method of digitizing an analog signal (such as a human voice) into a bit stream using simple analog to digital conversion techniques. [G.711] defines its use in the PSTN with two encoding laws, μ-law, used in N. America, and A-law, used elsewhere.
Quality of Service	Guarantees network bandwidth and availability for applications.
Real-time Transport Protocol	A protocol for encapsulating encoded voice and video streams. Refer to [RFC 3550].
Request for Comments	Technical policy documents approved by the IETF which are available on the World Wide Web at http://www.ietf.cnri.reston.va.us/rfc.html.
Session Initiation Protocol Plus	An extension to SIP.
Session Initiation Protocol	An application-layer control (signaling) protocol for creating, modifying, and terminating sessions with one or more participants.
Terminal Adapter	A device that converts an analog tip and ring interface into a digital signal; it includes a hybrid to convert from the 2-wire interface to 4-wire.
Upstream	The direction from the subscriber location toward the headend.
User Datagram Protocol	A connectionless protocol built upon Internet Protocol (IP).

4 ABBREVIATIONS AND ACRONYMS

This specification uses the following abbreviations:

ASO	Arbitrary Slice Ordering
AVC	Advanced Video Coding
CABAC	Context-Adaptive Binary Arithmetic Coding
CAVLC	Context-Based Adaptive Variable Length Coding
CIF	Common Intermediate Format
Codec	COder-DECoder
DQoS	Dynamic Quality of Service.
DTMF	Dual-tone Multi Frequency (tones)
ETSI	European Telecommunications Standards Institute
FMO	Flexible Macroblock Ordering
HD	High Definition
IETF	Internet Engineering Task Force
IP	Internet Protocol. An Internet network-layer protocol
ITU	International Telecommunication Union
ITU-T	International Telecommunications Union-Telecommunications Standardization Sector
MG	Media Gateway
MGCP	Media Gateway Control Protocol. Refer to [RFC 3435].
NCS	Network Call Signaling
PCM	Pulse Code Modulation
PLC	Packet Loss Concealment
PSTN	Public Switched Telephone Network
QCIF	Quarter Common Intermediate Format
QoS	Quality of Service
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
SIP	Session Initiation Protocol.
SIP+	Session Initiation Protocol Plus. An extension to SIP.
UDP	User Datagram Protocol. A connectionless protocol built upon Internet Protocol (IP).
VAD	Voice Activity Detection
VoIP	Voice over IP

5 COMMON CRITERIA FOR MEDIA TRANSPORT OVER IP

This section outlines the required basic functionality of the PacketCable architecture with respect to audio and video media handling. The key requirement for narrowband 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. PacketCable additionally provides the means for cable operators to offer superior audio communications quality, exceeding current PSTN standards, by using wideband audio codecs. Finally, PacketCable 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.

5.1 IP Network Criteria for Codec Support

5.1.1 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.

5.1.2 Latency Control

The ITU-T Recommendation [G.114], defines standards for network latency. Specifically, an end-to-end delay of 150ms 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:

- Audio or video sampling and analog-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
- 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 below.

5.1.2.1 Jitter Buffer Management

Jitter buffers are required to smooth out packet delay variation to provide for a continuous playout on a TDM or analog 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 pre-configured 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 playout 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.

5.1.2.2 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 30ms of audio (typically one, two, or, at most, three frames of compressed audio data. These packet sizes are only applicable to audio as the frame sizes for video 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 lockstep operation between the codec framing and packet generation.

5.1.2.3 Codec Selection

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

5.1.3 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 PacketCable device to support every possible codec technology that could be interconnected with a PacketCable system. Transcoding within the PacketCable network is inevitable. Transcoding is often associated with undesirable artifacts such as degraded voice quality and increased latency. However, the use of a high-quality and low delay codec mitigates degradation and delay build-up.

5.1.4 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, thus reducing the bandwidth required;
- Large packet sizes can 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. Regardless, this 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 encoding (VBR) occurs when a codec dynamically optimizes the compression bit stream to adapt the source data to network conditions as, for example, in cellular networks.

5.2 Overall Quality Targets

With respect to media quality, PacketCable-based services and applications should be designed with two related sets of "end-to-end" performance targets in mind. The first set of targets is derived from human factors and establishes metrics for performance that incorporate endpoints such as MTAs and IP phones. An example of such a metric is "mouth to ear delay" - i.e., from the mouth of the speaker to the ear of the listener.

The second set of performance targets relates to the underlying network. In this context the "ends" in "end-to-end" are the user-to-network interfaces (UNI), interpreted to include PSTN gateways. That is, "end-to-end" reduces to

"UNI-to-UNI" and explicitly excludes terminal equipment, even though those devices may participate in network performance monitoring.

With those distinctions in mind, following are the recommendations for both sets of target metrics for conversational services such as audio and video telephony. Targets for other services, such as audio and video streaming services, are not included in this version of the specification.

The recommended targets for the first, human factors-based set of metrics are as follows (see [G.109], [G.114], [G.131], [T1.508], [ETSI TR 102 493], and [ATSC]):

- One-way delay (e.g., mouth-to-ear or glass-to-glass): ≤ 150 ms[1]
- Skew between audio and associated video:
  - ≤ 30 ms audio advance over video[2]
  - ≤ 100 ms audio delay following video[3]
- Talker Echo Loudness Rating (TELR): ≥ 65 dB[4]
- E-model Transmission Rating Factor (R Factor): ≥ 80

The corresponding targets for the second, network-centric set of metrics are as follows, corresponding to a Class 0 network in Y.1541 (see [Y.1540], [Y.1541] and [G.1050]):

- One-way delay[5] : ≤ 100 ms
- Jitter[6]: ≤ 50 ms
- Packet loss ratio[7] : ≤ 0.1%

5.3 Media Security

There is no security defined for media transmission. For more details see [SEC TR].

[1] It is recognized that 150ms end-to-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 as for voice.
[2] Corresponding to approximately 1 NTSC video frame.
[3] Corresponding to approximately 3 NTSC video frames.
[4] A TELR of 65 dB corresponds to a Residual Echo Return Loss (RERL) of between 55 dB and 60 dB, depending on the type of line and telephone [G.100], [G.121], [T1.508].
[5] IP packet transfer delay (IPTD) in [Y.1540] and [Y.1541].
[6] IP packet delay variation (IPDV) in [[Y.1540] and [Y.1541]. Also note that, as far as the human user is concerned, jitter can only be perceived as gaps in the audio or video streams, resulting from lost or discarded packets.
[7] IP packet loss ratio (IPLR) in [Y.1540] and [Y.1541].

6 GAP ANALYSIS BETWEEN 3GPP, 3GPP2 AND PACKETCABLE CODEC REQUIREMENTS

This section 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 PacketCable codecs. In addition to the codec applications within the PacketCable 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 the IMS.

6.1 Audio Codecs

6.1.1 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 signaling 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 [TS 26.235] and [TS 24.229] recommends the use of DTMF [RFC 4733] 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 service and, especially, over the PSTN, which may thereby accelerate the substitution of the legacy wireline network.

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

Table 1 - Summary of 3GPP and 3GPP2 Audio Codecs

	3GPP	3GPP2
Narrowband Telephony	AMR	SMV EVRC
Wideband telephony	AMR-WB	VMR-WB

6.1.2 Audio Codec Analysis for PacketCable

One of the mandates of PacketCable 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, PacketCable codec selection includes additional codecs beyond those specified by 3GPP and 3GPP2. PacketCable also imposes strict requirements for the reliable transport of various types of voiceband data not encountered in cellular networks and therefore beyond the scope of 3GPP or 3GPP2. Fax, dial-up modem, point-of-sale terminal, and Telephone Devices for the Deaf (TDD) are all examples of voiceband data that are reliably supported by the PSTN and must be well supported by PacketCable networks.

6.2 Video Codecs

6.2.1 3GPP and 3GPP2 Video Codecs

3GPP and 3GPP2 specify mandatory and recommended codecs to allow cellular video telephony and streaming with QCIF resolution (176 x 144), consistent with a cellular handset's small screen size and cellular networks' limited bandwidth [TS 26.234] and [TS 26.235]. These requirements are summarized in Table 2.

Table 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.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.264 (recommended) Baseline Profile @ Level 1b, with constraint_set1_flag = 1 MPEG-4 Part 2 (recommended) Simple Profile @ Level 0b

6.2.2 Video Codec Analysis for PacketCable

PacketCable 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 resolutions suitable for cellular devices. PacketCable defines additional codecs and profiles to support multiple resolutions as is necessary to provide user satisfaction with both small-screen and larger-screen devices.

6.3 Quality Metrics

6.3.1 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.

6.3.2 Quality Metric Analysis for PacketCable

PacketCable 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. PacketCable 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 [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 PacketCable networks and peer IMS networks, including 3GPP or 3GPP2 cellular networks. Additionally applications based on PacketCable architecture may utilize VoIP metrics to resolve voice quality issues. Therefore, the support of the VoIP Metrics Block for conversational applications is defined for PacketCable UEs.

7 CODEC AND MEDIA REQUIREMENTS

This section specifies codec and RTP media requirements.

Figure 2 below shows an abstract architecture for the media flows in a PacketCable 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 PacketCable 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 PacketCable network and provides TDM trunk interfaces to the PSTN.

VoIP and video sessions between client devices in the PacketCable network and another VoIP network must traverse a Media Proxy, which acts as a gateway to protect the PacketCable network through 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 PacketCable network and a peer VoIP network.

VoIP or video sessions from client devices may also terminate at a Media Server within the PacketCable 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.

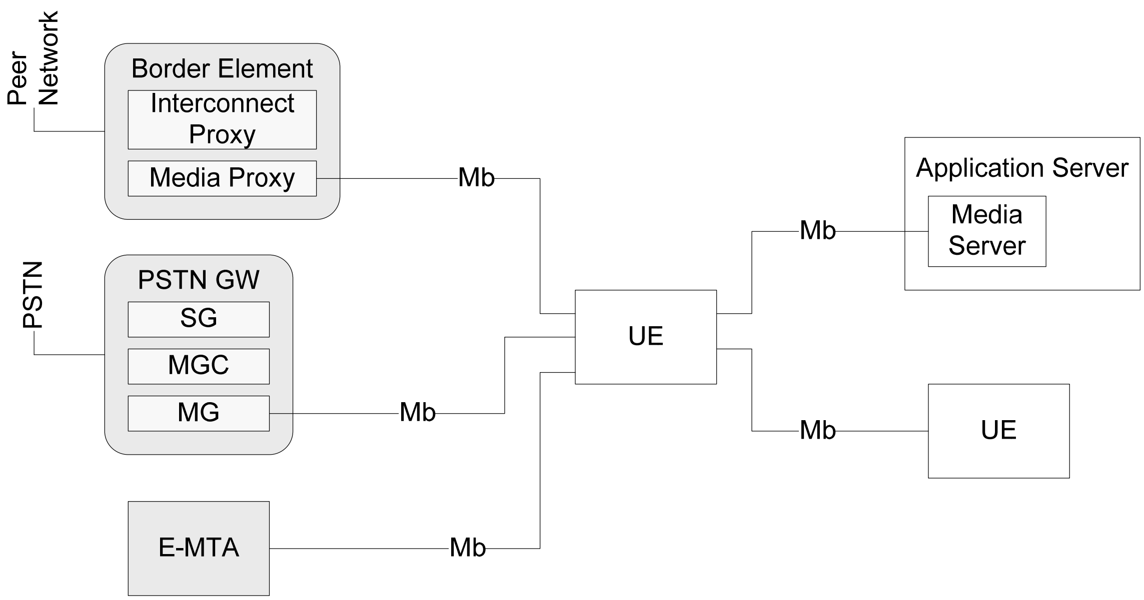

Figure 2 - RTP/RTCP Media Connection

The interfaces depicted in Figure 2 are described in Table 3.

Table 3 - Media Stream Interfaces

Reference Point	PacketCable Network Elements	Reference Point Description
Mb	UE - UE UE - MG UE - Border Element UE - AS 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 traveling across the Mb reference point can be audio traffic encoded by narrowband 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 2 but shown in Figure 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 section 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.

7.1 RTP Requirements

User Equipment and Media Gateways MUST support the Real-Time Transport Protocol (RTP) as defined in [RFC 3550] and [RFC 3551] for transport of audio and video media flows.

A Media Server that performs mixing of RTP streams MAY transmit contributing source lists (CSRC). This requirement is intended to allow mixers to omit CSRC lists, in compliance with [RFC 3550] and [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.

7.2 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.

7.2.1 General Requirements of the PacketCable RTCP Profile

User Equipment and Media Gateways MUST send and receive RTCP messages as described in [RFC 3550] and [RFC 3551] subject to the following over-riding requirements and clarifications.

Unless disabled via SDP exchange, 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 established once each endpoint has received a remote connection descriptor. Furthermore, unless disabled via SDP exchange, a PacketCable 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 non-essential bandwidth utilization, the IMS profile for RTP usage [TS 26.236] disables RTCP for wireless UEs while transmitting and receiving RTP packets. Sources of broadcast and multicast may also not wish to receive RTCP reports so as not to be inundated with RTCP packets sent from media recipients. Therefore, it is necessary to be able to disable RTCP transmission on a per-session basis. This is done via exchange of the SDP media-level bandwidth modifiers b=RS and b=RR defined in [RFC 3556]. [RFC 3556] specifies b=RS as the RTCP bandwidth for active data senders and b=RR as the RTCP bandwidth for recipients. If RS is set to 0 but RR is non zero, senders can still send RTCP packets with the restriction of the RR bandwidth.

For one-way broadcast and multicast applications, User Equipment and Media Gateways distributing RTP media MAY add the following line into the SDP media description to disable the return of RTCP receiver reports:

b=RR:0

For conversational services, User Equipment and Media Gateways MAY add the following line into the SDP media description to disable the exchange of RTCP packets:

b=RS:0

b=RR:0

User Equipment and Media Gateways receiving both these bandwidth parameters in the SDP set to zero, or, for receive-only streams, just the RR parameter set to zero, MUST not generate RTCP packets.

To avoid unnecessary network traffic, when RTCP is not disabled via SDP exchange, User Equipment or Media Gateways MAY stop sending RTCP packets to a remote endpoint if an ICMP port unreachable or another ICMP destination unreachable error (i.e., ICMP error type 3) is returned from the network for that RTCP destination.

To avoid unnecessary network traffic, User Equipment or Media Gateways MAY stop sending RTCP packets to a remote endpoint if no RTCP packets have been received within five (5) report transmission intervals. This requirement allows the endpoint to stop sending RTCP packets to 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 5s. 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 [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 sections 6.2 and 6.3 of [RFC 3550].

User Equipment and Media Gateways MUST receive RTCP messages, if sent by the remote communication peers. User Equipment and Media Gateways MUST NOT require them 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-PacketCable endpoints.

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

SSRC (Synchronization Source) 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:

SDES (Source Description): CNAME objects MUST NOT contain identity information (see definition below); CNAME field MUST be a cryptographically-random value generated by the endpoint in such a manner that endpoint identity is not compromised and MUST change on a per-session basis; NAME, EMAIL, PHONE, LOC objects SHOULD NOT be sent and, if sent, MUST NOT contain identity information. This requirement is intended to satisfy the requirements of [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.

SR (Sender Report): MUST be sent by User Equipment and Media Gateways transmitting RTP packets (as described in [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.

RR (Receiver Report): MUST be sent with report blocks if receiving but not sending RTP packets (as described in [RFC 3550]) and MUST be sent without report blocks if not sending or receiving RTP packets, except as previously described when errors occur or the remote endpoint does not send RTCP packets, in which case they MAY be sent.

APP (Application-Defined): MAY be sent as implementation needs dictate and MUST NOT contain identity info. User Equipment and Media Gateways MUST ignore and silently discard APP messages with unrecognized contents.

BYE (Goodbye): MUST be sent upon RTP connection deletion or when renegotiating SSRC upon collision detection and resolution (see below). 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 [RFC 3550]. This requirement is intended to ensure that all call progress conditions, such as on-hook notifications, are signaled using the higher-level signaling 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.

7.2.2 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.

7.2.3 Extended Statistics Reporting using RTCP-XR

The RTCP Extended Reports (XR) as defined in [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. PacketCable presently only defines use of the VoIP Metrics Report Block as described in Section 7.8.1, 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.

7.3 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.. This section describes the required specification for each 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 [RFC 4566]. In this case, the 0 represents a static payload type of μ-law PCM-coded, single channel audio, sampled at 8kHz. On the media attribute line (a), the first term defines the packet formation time (10ms).

Payload types other than those defined in [RFC 3550] are dynamically bound by using a dynamic payload type from the range 96-127, as defined in [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/sec.

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

7.3.1 SDP Use

7.3.1.1 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=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 lowercase, uppercase, and mixed upper/lowercase encodings of all attributes.

Note that SDP [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 [RFC 4733] 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 [RFC 4566].

Note: Refer to Sections 7.3.2, 7.4.2.4.1, and 7.4.2.6.2 for more specific information regarding the "fmtp" attribute.

sqn:
cdsc:
cpar:
cparmin:
cparmax:

As defined in [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 [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.

ptime:

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

Receive: This field is used to define the packetization interval for all codecs present in the media line of the SDP.

Note: [RFC 4566] defines the "maxptime" SDP attribute and [V.152] defines the "maxmptime" SDP attribute. The precedence of these attributes with respect to the "ptime" and "mptime" attributes is not defined at this time.

7.3.2 Session Description for audio/RED

The following SDP attributes are applicable to Audio Service Use for [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 lowercase, uppercase, and mixed upper/lowercase encodings of all attributes.

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

rtpmap:

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

Receive: 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 honor the recommendation in the answer and offer, respectively.

Following is an example of the media representation in SDP for describing a RFC 2198 primary and secondary encoding (one level of redundancy) involving G.729:

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

7.4 Narrowband Codec Specifications

Narrowband codecs are defined to operate on audio signals bandpass filtered to a frequency range of 300Hz - 3400 Hz [G.712] and sampled at 8000 samples/second. For codecs other than 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 G.711, the codec input is specified as 13 or 14-bit uniform PCM samples according to [G.711]. A comparison of the well-known narrowband codecs is provided in Annex C.

7.4.1 Supported Narrowband Codecs

The following sections describe every narrowband codec supported in PacketCable. Whether a particular narrowband 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 section MUST be met.

7.4.1.1 G.711

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/second. A G.711 encoder will create a 64 kbps bitstream. 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 narrowband audio communications.

G.711 (both μ-law and A-law versions) [G.711] MAY be supported by User Equipment and Media Gateways. User Equipment and Media Gateways MUST support 10ms, 20ms, and 30ms packetization rate when G.711 is used. It can be used to provide "fallback" for services such as fax, modem, and hearing-impaired services, as well as common gateway transcoding support.

7.4.1.1.1 Packet Loss Concealment

For G.711, User Equipment and Media Gateways SHOULD use the method defined in [G.711-I].

7.4.1.1.2 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 (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 [G.711-II].

7.4.1.1.3 Payload Header Format

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

7.4.1.1.4 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 [RFC 3551].

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
```

7.4.1.2 Internet Low Bit Codec (iLBC)

iLBC was selected by CableLabs as a codec standard suitable for packet-based communication networks. Additionally, iLBC has undergone IETF standardization process as a part of the IETF Audio Visual Transport (AVT) Working Group [RFC 3951], [RFC 3952]. Experimental track IETF RFC "internet Low Bit Rate Codec (iLBC)" [iLBC] contains the iLBC source code in floating point C.

iLBC MAY be supported by User Equipment and Media Gateways. iLBC provides two modes with coding rates of 13.3 kb/s and 15.2 kb/s using 30ms and 20ms frame sizes respectively. User Equipment and Media Gateways that implement iLBC MUST support both modes of operation.

7.4.1.2.1 Packet Loss Concealment

For iLBC, User Equipment and Media Gateways SHOULD use the method defined in [RFC 3951] for packet loss concealment.

7.4.1.2.2 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 (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 [G.711-II].

7.4.1.2.3 Payload Header Format

User Equipment and Media Gateways MUST support the payload header format as specified in [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 [RFC 3952] for iLBC. There are no options specific to this payload header format. The codec frame size mode (20ms or 30ms) is specified by out-of-band means.

7.4.1.2.4 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 [RFC 3951]).

If 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 30ms frame size).

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
```

Alternately, if 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 [RFC 3952].

7.4.1.3 Broad Voice 16 (BV16)

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

BV16 MAY be supported by User Equipment and Media Gateways. BV16 supports a coding rate of 16 kb/s with a frame size of 5ms. User Equipment and Media Gateways MUST support 10ms, 20ms, and 30ms packet sizes when BV16 is used.

7.4.1.3.1 Packet Loss Concealment

For BV16, User Equipment and Media Gateways SHOULD use the method defined in [BVOICE] for packet loss concealment.

7.4.1.3.2 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 (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 [G.711-II].

7.4.1.3.3 Payload Header Format

User Equipment and Media Gateways MUST support the payload header format specified in [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 [RFC 4298] for BV16. There are no options specific to this payload header format.

7.4.1.3.4 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 [RFC 4298]).

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
```

7.4.1.4 Adaptive Multi Rate (AMR)

The AMR codec [TS 26.090] was originally developed for use in GSM cellular systems by ETSI. It has now also been chosen for use in 3G cellular systems by 3GPP. It is also a mandatory codec in the 3GPP IP Multimedia Subsystem (IMS) specifications [TS 26.235]. PacketCable has a mandate to provide interworking to cellular systems. Recommending the use of AMR guarantees end-to-end narrowband codec interoperability between User Equipment or Media Gateways and 3GPP cellular networks.

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 kb/s. Three of these encoding modes already exist as independent coding standards: the 12.2 kb/s mode as [TS 46.060], the 7.4 kb/s mode as IS-641 [ANSI 136-C], and the 6.7 kb/s mode as PDC-EFR [ARIB 27H]. All encoding modes use a standard 20ms 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.

7.4.1.4.1 Packet Loss Concealment

User Equipment and Media Gateways supporting AMR SHOULD use the method defined in [TS 26.091] for packet loss concealment.

7.4.1.4.2 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 a PacketCable 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 [TS 26.093], and [TS 26.094]. The choice of VAD1 or VAD2 as outlined in [TS 26.094] is left to the vendor as this choice does not require any signaling and does not impact on interworking.
- support generation and handling of SID frames in accordance with [TS 26.093] and [TS 26.101]. (The use of extra SID frame types in [TS 26.093], i.e., GSM-EFR SID, TDMA-EFR SID, and PDC-EFR SID are NOT required.)
- support comfort noise generation in accordance with [TS 26.092].

7.4.1.4.3 Payload Header Format

The payload header format for AMR is specified in [RFC 3267]. This RFC outlines a range of supported features and options. A profile of [RFC 3267] outlining the options supported in IMS applications is given in [TS 26.236]. User Equipment and Media Gateways supporting AMR MUST support the Payload Header format specified in

[RFC 3267] with the options specified in [TS 26.236]. 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 signaled 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 [TS 26.236]. User Equipment and Media Gateways supporting AMR encode and decode implementations MAY support octet-aligned mode.
- CMR (Codec Mode Request): 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 kb/s and AMR SID frames), as well as switching to any mode at any 20ms frame boundary. The codec mode a near-end AMR decoder prefers to receive is signaled 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 bi-directional 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 [RFC 3267]. The use of CMR itself does not require out-of-band signaling.

  In certain transport networks, the full range of codec modes supported may be restricted to a defined subset. For example, 3GPP usage specified in [TS 44.018] describes an Active Codec Mode Set of up to 4 codec modes to be used on a particular call. The signaling 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 neighboring modes in the active codec mode set can be switched to, are signaled 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 [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 [TS 26.236].

- Redundant Transmission: The RTP payload format specified in [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 [RFC 2198]. The use of redundant transmission does not require out-of-band signaling. 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 [TS 26.236], 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. [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 signaling 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 [TS 26.236], frame interleaving MUST NOT be used in User Equipment and Media Gateway AMR implementations.
- Frame CRCs: [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 signaling the 'crc' parameter out-of-band. In accordance with [TS 26.236], frame CRCs MUST NOT be used in User Equipment and Media Gateway AMR implementations.

- Robust Sorting: If multiple AMR encodings are packed into one [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 signaling the 'robust-sorting' parameter out-of-band. In accordance with [TS 26.236], simple sorting MUST be supported in User Equipment and Media Gateway AMR implementations and robust sorting MUST NOT be used.

7.4.1.4.4 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 [RFC 3267]). When conveying information by SDP, the maxptime attribute MUST be present and sent with a value of 20msec.

Following is an example of the media representation in SDP for describing AMR when bandwidth-efficient mode is used with only 4 codec modes (4.75, 7.4, 10.2, 12.2 kb/s) in the active codec mode set and switching between adjacent active modes permitted only if switches are at least 2 frame blocks apart:

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

7.4.1.5 Selectable Mode Vocoder (SMV)

The SMV codec [TS C.S0030] 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. PacketCable has a mandate to provide interworking to cellular systems. Recommending the use of SMV guarantees end-to-end narrowband 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 4 possible bit rates: 0.8, 2.0, 4.0, and 8.55 kb/s. All bit rates use a standard 20ms 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% Voice Activity Factor), the average bit rate for SMV in Mode 0 (highest quality) is 3.70 kb/s while for Mode 5 (lowest quality) it is 1.85 kb/s. The configuration of operating mode is done externally to the codec and is out of scope of this document. 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.

7.4.1.5.1 Packet Loss Concealment

User Equipment and Media Gateways supporting SMV SHOULD use the frame erasure concealment method defined in [TS C.S0030] for packet loss concealment.

7.4.1.5.2 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 [TS C.S0030]. The decision to use VAD option A or VAD option B is left to the vendor as this choice does not require any signaling and does not impact on interworking. No Discontinuous Transmission (DTX), Silence Insertion Descriptors (SID), or Comfort Noise Generation schemes are currently defined for SMV as SMV is designed for continuous transmission.

7.4.1.5.3 Payload Header Format

User Equipment and Media Gateways supporting SMV MUST use the payload header format as specified in [RFC 3558]. This RFC outlines a range of supported features and options. Similar to 3GPP, 3GPP2 has not provided a profile for this RFC.

7.4.1.5.4 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 [RFC 3558]).

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=ptime: 20
a=fmtp: 97 maxinterleave=3
```

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

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

7.4.1.6 Enhanced Variable-Rate codec (EVRC)

The EVRC codec [TS C.S0014] 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. PacketCable has a mandate to provide interworking to cellular systems. Recommending the use of EVRC guarantees end-to-end narrowband 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 3 possible bit rates: 0.8, 4.0, and 8.55 kb/s. All bit rates use a standard 20ms frame size. If EVRC is supported, all encoding rates MUST be supported.

7.4.1.6.1 Packet Loss Concealment

User Equipment and Media Gateways supporting EVRC SHOULD use the frame erasure concealment method defined in [TS C.S0014].

7.4.1.6.2 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 (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 [G.711-II].

7.4.1.6.3 Payload Header Format

User Equipment and Media Gateways MUST use the payload header format for EVRC as specified in [RFC 3558]. This RFC outlines a range of supported features and options. Similar to 3GPP, 3GPP2 has not provided a profile for this RFC.

7.4.1.6.4 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 [RFC 3558]).

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=ptime: 20
a=fmtp: 98 maxinterleave=3
```

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

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

7.4.2 Feature Support

The ability to offer a competitive telephone service depends on more than the use of toll-quality codecs. The end-to-end narrowband 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 a PacketCable 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 behavior rather than with packet networks. This means that the PacketCable 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, analog modem, and hearing-impaired TTY devices, and DTMF digits.

7.4.2.1 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 milliseconds, talker echo can become discernible. An echo canceller is used to remove this echo.

User Equipment and Media Gateways that provide analog 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 [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 (8ms or less). For PSTN media gateway applications, however, User Equipment and Media Gateways MUST support echo canceller lengths of 48ms minimum consistent with [T1.508]. Vendors MAY choose to differentiate their products by providing longer echo canceller lengths suitable for their application, 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.

7.4.2.2 Asymmetrical Services Support

User Equipment and Media Gateways SHOULD be capable of supporting different codecs for upstream and downstream audio channels. This allows potential optimization of device resources, network bandwidth, and user service quality.

7.4.2.3 Hearing-impaired Services Support

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 G.711 codec is recommended. Note that no detection or switching is required for cellular text modems, which are inherently compatible with AMR and other cellular codecs.

7.4.2.4 DTMF Relay

RTP payload formats and usage to carry DTMF are specified in [RFC 4733] and call progress tones across an IP network either as recognized "telephone-events" or as a set of parameters defining a tone by its volume, frequency, modulation and duration of its components. Besides the transport of DTMF tones across an IP network, [RFC 4733] also allows for the remote collection of DTMF digits by a media gateway to relieve an Internet end system (e.g., media server) of having to do this. Other advantages of [RFC 4733] include inherent redundancy to cope with packet loss and the means to allow IP phones to generate DTMF digits when signaling to the PSTN without requiring DTMF senders. The requirements for DTMF relay for User Equipment and Media Gateways in this section 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 DTMF telephone-events 0-15. Other telephone-events MAY be supported. Negotiated events MUST be transferred via [RFC 4733] telephone-event packets regardless of the codec specified for the speech.

[RFC 4733] references ITU-T Q.24 in defining the minimum DTMF tone duration of 40 ms. Additionally, ITU-T Q.24 includes a duration range lower than 40 ms when the DTMF tones may be accepted as DTMF digits (as low as 20 ms). For North American networks, LSSGR [GR 506] specifies that tone durations greater than 40ms 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 40ms. Total on-off cycle times of 93ms are to be accepted, but 100ms is to be generated as both minimum and objective. Where DTMF tones are used for text telephony, [RFC 4733] references ITU-T V.18 Annex B, and recommends the DTMF detector at a sending gateway recognize tones and pauses of ≥40 ms. It further recommends that short-duration tones be indicated as lasting a minimum of 70ms and pauses a minimum of 50ms consistent with a V.18 Annex B DTMF sender.

Considering these industry requirements, User Equipment or Media Gateways with analog or Time Division Multiplexing (TDM) interfaces MUST detect DTMF tones of 40ms or more and report their duration relative to the RTP timestamp. User Equipment and Media Gateways MAY detect DTMF digits of duration greater than 23ms, but User Equipment or Media Gateways MUST NOT report DTMF digits when their duration is less than 23ms. User Equipment and Media Gateways that originate DTMF telephone-events MUST specify a minimum of 50ms 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. User Equipment and Media Gateways detecting DTMF tones in environments where V.18 Annex B TTY devices are used MAY enforce a minimum tone duration indication of 70ms and an inter-digit time of 50ms.

The packetization time of telephone-event packets in the transmit direction MUST be equal to the same packetization time as the selected audio codec. Therefore, the packetization time of telephone-event packets has the same range of intervals, i.e., 10, 20, and 30ms.

[RFC 4733] allows the detecting gateway to either replace the audio when transmitting telephone-event packets or send both audio and telephone-events concurrently using RFC 2198 redundancy headers. To avoid increasing bandwidth requirements, User Equipment and Media Gateways originating DTMF telephone-event packets MUST stop sending audio whenever a DTMF digit is detected, with suppression of audio continuing until retransmission of the end-of-event packets is complete. An audio packet being constructed at the time of a DTMF detection should be discarded.

In accordance with [RFC 4733], 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. Audio packets being replaced by telephone-event 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 (with the E-Bit flag set), 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. [RFC 4733] 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.

When a minimum DTMF tone playout duration is configured, User Equipment and Media Gateways MUST play out DTMF tones for the greater of the duration specified in the [RFC 4733] end-of-event packet or the configured minimum tone duration. 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 [RFC 4733] 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 45ms inter-digit time (silence) following play-out of the DTMF tone.

7.4.2.4.1 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 lowercase, uppercase, and mixed upper/lowercase encodings of all attributes.

Note that SDP [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 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.

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
```

7.4.2.5 Fax and Modem Support

User Equipment and Media Gateways may need to support analog 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 analog 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 G.711 for the remainder of the session unless fax relay is to be used as specified in Section 7.4.2.6 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 standard [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 [V.152] and specified in this document in Section 7.4.2.7. 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 V.152 with [RFC 2198] redundancy also makes the transmission more resilient to packet loss in the network. This is an important feature for V.152 since packet loss causes modems to drop in speed or disconnect. User Equipment and Media Gateways MAY support V.152 with [RFC 2198] redundancy as defined in this specification.

7.4.2.6 Fax Relay

PacketCable 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 [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 T.38 MUST support version 3 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 0 with which all implementations are required to interoperate. In accordance with version 3, User Equipment and Media Gateways MUST therefore, support the V.27ter, V.29, V.17 modem protocols for page transmission at transfer rates up to 14.4kbps and the V.34 modem protocol for page transmission at transfer rates up to 33.6kbps.

T.38 allows TCP, UDPTL, and RTP as alternative transport protocols. REQ16632 PacketCable UEs and Media Gateways MUST support UDPTL and RTP with the choice between these being a function of CSCF or application server preference and the end-to-end codec negotiation at initial call establishment. Negotiating T.38 as an RTP audio codec requires a reservation of DSP resources for T.38 even for voice calls but this allows for autonomous switching between voice and T.38 fax relay at the UE and media gateway without CSCF or application server involvement. Using RTP also has the benefit of allowing for PacketCable media security. REQ16633 If T.38 over RTP is negotiated at call setup time UEs and Media Gateways MUST autonomously switch to T.38 fax relay upon fax detection using RTP as the transport protocol. If T.38 with RTP is negotiated at call setup, and RTP T.38 packets are received, User Equipment and Media Gateways MUST immediately transition to sending T.38 packets. Within 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 T.38. When using redundancy with UDPTL, User Equipment and Media Gateways MUST transmit T.30 control message data with a redundancy level of 4 and T.4 phase C data with a redundancy level of 1 subject to this fitting within the maximum datagram size allowed by the remote T.38 device. If a redundancy level of 1 for the T.4 phase C data does not fit within the maximum datagram size allowed by the remote T.38 device, User Equipment and Media Gateways MUST transmit the T.4 phase C data without redundancy. In the receive direction, User Equipment and Media Gateways MUST also support a T.30 control message data with a redundancy level of 4 and T.4 phase C data with a redundancy level of 1 but SHOULD be prepared to receive any redundancy in UDPTL that fits within the maximum datagram size that the User Equipment and Media Gateway declares. Redundancy with RTP is based on [RFC 2198] while FEC is based on [RFC 2733]. When using T.38 with RTP, User Equipment and Media Gateways MUST support redundancy levels of 1 and 2. User Equipment and Media Gateways MUST send T.38 with RTP using either one or two levels of redundancy consistent with the T.38/RTP and RED/8000 negotiation. User Equipment and Media Gateways MUST not send T.38 with RTP with redundancy if the remote SDP does not contain a dynamic payload type mapped to RED/8000.

T.38 does not currently define any security authentication or privacy mechanisms for UDPTL. Consequently, T.38 sessions using UDPTL will not have secure media at the transport level. T.38 must be carried over RTP when secure fax relay is required; see Section 5.3.

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

For Group 3 rates, User Equipment and Media Gateways MUST be prepared to receive a T.38 UDPTL fax packet of up to 152. This is based on 40ms packetization period and a 14.4kbps 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 33.6kb/s, User Equipment and Media Gateways MUST be prepared to receive a T.38 UDPTL fax packet of up to 346 bytes for the same 40ms packetization period.

For QoS and NAT traversal considerations, T.38 fax transmission packets over RTP MUST use the same port used by the voice packets for the connection. Upon transition to UDPTL T38 via a new offer/answer exchange, User Equipment and Media Gateways SHOULD immediately send T.38 "No signal" indicator packets if the User Equipment or Media Gateway would not otherwise be sending signal or data packets. This is to maintain or establish NAT bindings. In addition, User Equipment and Media Gateways MUST send T.38 fax packets at a default 20ms packetization period in the upstream unless another packetization period is negotiated (10/20/30ms).

Table 5 shows the flowspec parameters for 10/20/30ms 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.

7.4.2.6.1 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 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 lowercase, uppercase, and mixed upper/lowercase encodings of all attributes.

SDP [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 [T.38]: The recipient of the offer MUST accept that 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 [T.38]: Early implementations of 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 the T.38 specification. 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 be in the range of 152 to 160 bytes for a 14.4 kb/s Group 3 transmission and 346 to 360 bytes for a 33.6kb/s Super Group 3 transmission to allow a remote T.38 device to transmit using a 40ms packetization period and a redundancy level of 1. The value indicated 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 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 supposed 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).

7.4.2.6.2 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 lowercase, uppercase, and mixed upper/lowercase encodings of all attributes.

Note that SDP [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 [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 following "audio/t38" format specific parameters:

T38FaxVersion:

As defined in [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 [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 the T.38 specification. 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. The value indicated MUST be sufficient to accommodate the T.38 Internet Fax Packet (IFP) at the allowed maximum data rate and packetization time, the RFC 2198 redundancy header, and the redundant IFP packets. For T.38 carried over RTP, the T38FaxMaxDatagram value is defined as the RTP payload.

Receive: Endpoints MUST NOT send a T.38 RTP payload larger than that specified in the T38FaxMaxDatagram attribute. As part of codec negotiation, and prior to sending any T.38 data, the endpoint MUST ensure that T.38 transmission can adhere to the limits defined by this attribute. If the specified T38FaxMaxDatagram value is too small to support redundancy, 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 at a specified packetization time, the endpoint MUST NOT send a 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.

7.4.2.7 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 [V.152] along with [RFC 2198] redundancy. 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. 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. VBD transmission MUST be of the same packetization time as the selected audio codec when a voice call is switched to VBD. 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. If V.152 and T.38 are both enabled, T.38 MUST be preferred over V.152.

Table 5 shows the flowspec parameters for 10/20/30ms 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.

7.4.2.7.1 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 lowercase, uppercase, and mixed upper/lowercase encodings of all attributes.

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

gpmd:

The "gpmd" attribute is applied at the media level as defined in [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 [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 [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 [V.152]. This does not preclude support for voiceband data transmission via G.711 if G.711 is negotiated as an audio media type.

pmft:

The "pmft" attribute is applied at the session level as defined in [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 4734] 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 2100Hz Answer tones (ANS, /ANS, ANSam, and /ANSam) that a receiver can handle. If an implementation chooses use of [RFC 4734] 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 4734] telephone events.

7.4.3 Codec Naming and Flow Spec Parameters for Narrowband Codecs

Narrowband codecs defined in this specification MUST be encoded with the following string names in the rtpmap parameter:

Table 4 - Narrowband 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 4733] DTMF	telephone-event	Telephone-event/8000
T.38 using RTP	T38	T38/8000

Unknown rtpmap parameters SHOULD be ignored if they are received.

For every defined codec, whether it is represented in SDP as a static or dynamic payload type, Table 5 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 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 PacketCable security considerations.

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

b= <modifier> : <bandwidth-value>

For example:

b= AS:99

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

Table 5 - Mapping of Narrowband Audio Codec Session Description Parameters to Flowspec

Parameters from Session Description			**Flowspec parameters**		**Comments**
RTP/AVP code	**rtpmap**	**ptime (msec)**	**Values b,m,M[1]**	**Values r,p[2]**	
0	<none>	10	120 bytes	12,000 bytes/sec	G.711 A-law using the Payload Type defined by IETF
0	<none>	20	200 bytes	10,000 bytes/sec	
0	<none>	30	280 bytes	9,334 bytes/sec	
96-127	PCMU/8000	10	120 bytes	12,000 bytes/sec	G.711 A-law PCM, 64 kb/sec, default codec
96-127	PCMU/8000	20	200 bytes	10,000 bytes/sec	
96-127	PCMU/8000	30	280 bytes	9,334 bytes/sec	
8	<none>	10	120 bytes	12,000 bytes/sec	G.711 A-law using the Payload Type defined by IETF
8	<none>	20	200 bytes	10,000 bytes/sec	
8	<none>	30	280 bytes	9,334 bytes/sec	
96-127	PCMA/8000	10	120 bytes	12,000 bytes/sec	G.711 A-law PCM, 64 kb/sec, default Codec
96-127	PCMA/8000	20	200 bytes	10,000 bytes/sec	
96-127	PCMA/8000	30	280 bytes	9,334 bytes/sec	
96-127	iLBC/8000	20	78 bytes	3900 bytes/sec	iLBC, FB-LPC, 15.2 kb/s, 20 ms frame size with 5 ms lookahead; 13.3 kb/s, 30 ms frame with 10 ms lookahead
96-127	iLBC/8000	30	90 bytes	3000 bytes/sec	
96-127	BV16/8000	10	60 bytes	6,000 bytes/sec	BV16 (narrow-band), 16kb/sec
96-127	BV16/8000	20	80 bytes	4,000 bytes/sec	
96-127	BV16/8000	30	100 bytes	3,334 bytes/sec	
96-127	AMR/8000	20	54 bytes	2,700 bytes/sec	AMR at 4.75kb/s
96-127	AMR/8000	20	55 bytes	2,750 bytes/sec	AMR at 5.15kb/s
96-127	AMR/8000	20	56 bytes	2,850 bytes/sec	AMR at 5.9kb/s
96-127	AMR/8000	20	58 bytes	2,950 bytes/sec	AMR at 6.7kb/s
96-127	AMR/8000	20	60 bytes	3,000 bytes/sec	AMR at 7.4kb/s
96-127	AMR/8000	20	62 bytes	3,100 bytes/sec	AMR at 7.95kb/s
96-127	AMR/8000	20	67 bytes	3,350 bytes/sec	AMR at 10.2kb/s
96-127	AMR/8000	20	72 bytes	3,600 bytes/sec	AMR at 12.2kb/s
96-127	SMV0/8000[3]	20	42 bytes	2,100 bytes/sec	SMV at 0.8 kb/s
96-127	SMV0/8000[3]	20	45 bytes	2,250 bytes/sec	SMV at 2.0 kb/s
96-127	SMV0/8000[3]	20	50 bytes	2,500 bytes/sec	SMV at 4.0 kb/s
96-127	SMV0/8000[3]	20	62 bytes	3,100 bytes/sec	SMV at 8.55 kb/s
96-127	EVRC0/8000[3]	20	42 bytes	2,100 bytes/sec	EVRC at 0.8 kb/s
96-127	EVRC0/8000[3]	20	50 bytes	2,500 bytes/sec	EVRC at 4.0 kb/s
96-127	EVRC0/8000[3]	20	62 bytes	3,100 bytes/sec	EVRC at 8.55 kb/s
96-127	red/8000	10	205 bytes	20,500 bytes/sec	RFC 2198 redundancy used for G.711 used as a V.152 codec with redundancy of level 1
96-127	red/8000	20	365 bytes	18,250 bytes/sec	
96-127	red/8000	30	525 bytes	17,500 bytes/sec	

Parameters from Session Description			Flowspec parameters		Comments
RTP/AVP code	rtpmap	ptime (msec)	Values b,m,M[1]	Values r,p[2]	
96-127	T38/8000	10	60	6,000 bytes/sec	T.38 RTP Group 3 fax relay no redundancy
96-127	T38/8000	20	78	3,900 bytes/sec	
96-127	T38/8000	30	96	3,200 bytes/sec	
96-127	T38/8000	10	84	8,400 bytes/sec	T.38 RTP V.34 fax relay no redundancy
96-127	T38/8000	20	126	6,300 bytes/sec	
96-127	T38/8000	30	168	5,600 bytes/sec	
96-127	T38/8000	10	85	8,500 bytes/sec	T.38 RTP Group 3 fax relay redundancy level 1
96-127	T38/8000	20	121	6,050 bytes/sec	
96-127	T38/8000	30	157	5,234 bytes/sec	
96-127	T38/8000	10	133	13,300 bytes/sec	T.38 RTP V.34 fax relay redundancy level 1
96-127	T38/8000	20	217	10,850 bytes/sec	
96-127	T38/8000	30	309	10,300 bytes/sec	
96-127	T38/8000	10	121	12,100 bytes/sec	T.38 RTP Group 3 fax relay redundancy level 2
96-127	T38/8000	20	175	8,750 bytes/sec	
96-127	T38/8000	30	217	7,234 bytes/sec	
96-127	T38/8000	10	181	18,100 bytes/sec	T.38 RTP V.34 fax relay redundancy level 2
96-127	T38/8000	20	319	15,950 bytes/sec	
96-127	T38/8000	30	445	14,834 bytes/sec	
N/A	N/A	10	50 bytes	5,000 bytes/sec	T.38 UDPTL Group 3 fax relay (without redundancy
N/A	N/A	20	68 bytes	3,400 bytes/sec	
N/A	N/A	30	86 bytes	2,867 bytes/sec	
N/A	N/A	10	74 bytes	7,400 bytes/sec	T.38 UDPTL V.34 fax relay (without redundancy)
N/A	N/A	20	116 bytes	5,800 bytes/sec	
N/A	N/A	30	158 bytes	5,267 bytes/sec	
N/A	N/A	10	72 bytes	7,200 bytes/sec	T.38 UDPTL Group 3 fax relay (with T.4 redundancy level 1)
N/A	N/A	20	108 bytes	5,400 bytes/sec	
N/A	N/A	30	144 bytes	4,800 bytes/sec	
N/A	N/A	10	120 bytes	12,000 bytes/sec	T.38 UDPTL V.34 fax relay (with T.4 redundancy level 1)
N/A	N/A	20	204 bytes	10,200 bytes/sec	
N/A	N/A	30	288 bytes	9,600 bytes/sec	

TABLE NOTES:

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

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

[3] Header-Free payload header format is assumed for the SMV and EVRC codecs.

7.5 Wideband Codec Specification

For the purpose of this specification, wideband codecs are defined as those which operate on audio signals band-pass filtered to a frequency range of 50 Hz - 7 kHz and sampled at 16,000 samples/second. Similar to narrowband, 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 Annex D.

7.5.1 Supported Wideband Codecs

The following sections describe every wideband codec supported in PacketCable. 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 section MUST be met.

7.5.1.1 G.722

G.722 [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. User Equipment and Media Gateways MUST support 10ms, 20ms, and 30ms packetization rates when G.722 is used. G.722 is a multi-rate wideband speech codec for 16 kHz sampled signals. It has three selectable bit rates: 48, 56, and 64 kb/s. The 48 kb/s version of G.722 produces medium-quality wideband speech, and the 56 kb/s and 64 kb/s versions produce good- to high-quality wideband speech. User Equipment and Media Gateways using the G.722 codec MUST support 64kb/s and SHOULD support 56 and 48kb/s.

7.5.1.1.1 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.

7.5.1.1.2 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 (Discontinuous Transmission - DTX) to reduce bandwidth using a mechanism of the vendor's choice. If silence suppression is used with G.722, User Equipment and Media Gateways SHOULD transmit Silence Insertion Descriptor frames as specified in [G.711-II].

7.5.1.1.3 Payload Header Format

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

7.5.1.1.4 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 [RFC 3551].

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

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

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

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

7.5.1.2 Broad Voice 32 (BV32)

BroadVoice32 MAY be supported in User Equipment and Media Gateways. User Equipment and Media Gateways MUST support 10ms, 20ms, and 30ms packetization rates when BroadVoice32 is used. BroadVoice32 (BV32) is a wideband speech codec for 16 kHz sampled signals. BV32 is a 32 kb/s, wideband speech codec. BV32 is very similar to BV16 in terms of the coding algorithm.

7.5.1.2.1 Packet Loss Concealment

BV32 has an associated PLC mechanism similar to BV16. User Equipment and Media Gateways SHOULD use the method defined for BV32.

7.5.1.2.2 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 (Discontinuous Transmission - 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 [G.711-II].

7.5.1.2.3 Payload Header Format

User Equipment and Media Gateways MUST support the payload header format for BV32 as specified in [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 [RFC 4298]. There are no options specific to this payload header format.

7.5.1.2.4 Session Description

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

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=ptime: 20
```

7.5.1.3 Adaptive Multi Rate - Wideband (AMR-WB/G.722.2)

The AMR-WB codec [TS 26.190] was originally developed for use in GSM cellular systems by ETSI. It has also been standardized by ITU-T as G.722.2 [G.722]. 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 [TS 26.235]. PacketCable has a mandate 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 [ETSI TR 102 493] for more details.

AMR-WB is a variable-bit-rate wideband speech codec for 16 kHz sampled signals. It has 9 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 kb/s. Except for the lowest two modes of 6.60 and 8.85 kb/s, AMR-WB gives good to high speech quality 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 20ms 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.

7.5.1.3.1 Packet Loss Concealment

User Equipment and Media Gateways supporting AMR-WB codec SHOULD use the method defined in [TS 26.191] for packet loss concealment.

7.5.1.3.2 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 [TS 26.193] and [TS 26.194].

- support generation and handling of SID frames in accordance with [TS 26.191], [TS 26.193] and [TS 26.201].
- support comfort noise generation in accordance with [TS 26.192] .

7.5.1.3.3 Payload Header Format

The payload header format is specified in [RFC 3267]. This RFC outlines a range of supported features and options. User Equipment and Media Gateways MUST adhere to the formats specified in [RFC 3267]. A profile of RFC3267 outlining the options supported in IMS applications is given in the 3GPP specification [TS 26.236] 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 signaled 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 [TS 26.236]. User Equipment and Media Gateways supporting AMR-WB encode and decode implementations MAY support octet-aligned mode.
- CMR (Codec Mode Request): 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 kb/s and AMR-WB SID frames) as well as switching to any mode at any 20ms frame boundary. The codec mode a near-end AMR-WB decoder prefers to receive is signaled 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 bi-directional 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 [RFC 3267]. The use of CMR itself does not require out-of-band signaling.

In certain transport networks, the full range of codec modes supported may be restricted to a defined subset. For example, 3GPP usage specified in [TS 44.018] describes an Active Codec Mode Set of up to 4 codec modes to be used on a particular call. The signaling 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 neighboring modes in the active codec mode set can be switched to, are signaled 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 [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 [TS 26.236].

- Redundant Transmission. The RTP payload format specified in [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 [RFC 2198] The use of redundant transmission does not require out-of-band signaling. 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 [TS 26.236], 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. [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 signaling 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 [TS 26.236], frame interleaving MUST NOT be used in AMR-WB implementations.

- Frame CRCs: [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 [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 signaling the 'crc' parameter out-of-band. In accordance with [TS 26.236], frame CRCs MUST NOT be used in AMR-WB implementations.
- Robust Sorting: If multiple AMR-WB encodings are packed into one [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 signaling the 'robust-sorting' parameter out-of-band. In accordance with [TS 26.236], simple sorting MUST be supported in AMR-WB implementations and robust sorting MUST NOT be used.

7.5.1.3.4 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 [RFC 3267]). When conveying information by SDP, the maxptime attribute MUST be present and sent with a value of 20msec.

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

7.5.1.4 Variable Rate Multi-Mode - Wideband (VMR-WB)

VMR-WB [TS C.S0052-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. PacketCable has a mandate 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 5 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, 3 communicate at 13.3, 6.2, 2.7 and 1.0 kb/s. Mode 4 communicates at 8.55, 4.0, 0.8 kb/s. The VMR-WB modes are determined by the Service Option used within the CDMA2000 network. For Service Option 62, modes 0, 1, 2 are supported. Service Option 63 supports mode 4. In addition, mode 3 allows interoperable operation with AMR-WB. However, only Service Options 62, 63 are specified for CDMA2000 terminals that support VMR-WB. All bit rates use a standard 20ms frame size. User Equipment and Media Gateways supporting VMR-WB MUST support all encoding rates within Service Options 62, 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 kb/s while for Modes 1 and 2 it is 7.6 kb/s and 6.2 kb/s respectively. The configuration of operating mode is done externally to the codec and is out of scope of this document. 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

7.5.1.4.1 Packet Loss Concealment

User Equipment and Media Gateways implementing VMR-WB SHOULD use the method defined in VMR-WB specification for packet loss concealment.

7.5.1.4.2 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 a PacketCable UE to handle SID frames and generate CNG in the same fashion as a 3GPP2 cellular device.

7.5.1.4.3 Payload Header Formats

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

7.5.1.4.4 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 [TS C.S0052-A]).

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 [TS C.S0052-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 1, 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-align mode of operation (not bandwidth efficient) and only codec modes 0, 1, and 2 of the AMR-WB codec must be used (6.6 kbps, 8.85 kbps, and 12.65 kbps, respectively). An example of such an offer-answer exchange between a CDMA2000 and a WCDMA terminal can be found in [TS C.S0052-A].

7.5.2 Feature Support

Unlike narrowband 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.

7.5.2.1 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 analog POTS interfaces or PSTN interfaces are not expected to allow use of wideband codecs on connections that use these interfaces.

7.5.2.2 Echo Cancellation Support

Wideband audio terminals are inherently 4-wire with separate transmit and receive signal directions. As such, the traditional 2-wire to 4-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 narrowband codecs for which echo cancellation requirements are specified in Section 7.4.2.1.

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.

7.5.2.3 Asymmetrical Services Support

The requirement specified in Section 7.4.2.2 for User Equipment and Media Gateways to support different codecs for upstream and downstream audio channels also applies to wideband codecs.

7.5.2.4 Hearing-impaired Services Support

Acoustically-coupled text telephone devices may be used with wideband codecs. Therefore, wideband User Equipment and Media Gateways MUST support detection of [V.18] Annex A hearing-impaired tones in the same way as described in Section 7.4.2.3.

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. The G.711 codec is recommended. 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.

7.5.2.5 DTMF Relay

Though a legacy PSTN subscriber signaling 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 narrowband to wideband. DTMF relay is specified in Section 7.4.2.4 for this purpose and is applicable to wideband codecs as well as narrowband codecs.

User Equipment and Media Gateways supporting wideband codecs MUST relay DTMF digits using [RFC 4733] telephone-event packets per the requirements specified in Section 7.4.2.4. 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 62.5µs, corresponding to 16,000 samples per second. In this case the SDP used to negotiate DTMF relay for a wideband session MUST specify Telephone-event/16000 as a media attribute.

7.5.2.6 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 analog 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.

7.5.3 Codec Naming and Flow Spec Parameters for Wideband Codecs

The wideband codecs defined in this specification MUST be encoded with the following string names as defined in Table 6 in the rtpmap parameter:

Table 6 - Wideband Audio Codec rtpmap Parameters

Codec	Literal Codec Name	rtpmap Parameter
BroadVoice32	BV32	BV32/16000
G.722 at 48 kb/s	G722-48	G722-48/8000
G.722 at 56 kb/s	G722-56	G722-56/8000
G.722 at 64 kb/s	G722-64	G722-64/8000
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 7 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 7 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 PacketCable security considerations.

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

b= <modifier> : <bandwidth-value>

For example:

b= AS:99

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

Table 7 - Mapping of Wideband Audio Codec Session Description Parameters to Flowspec

Parameters from Session Description			Flowspec parameters		Comments
RTP/AVP code	rtpmap	ptime (msec)	Values b,m,M[1]	Values r,p[2]	
96-127	BV32/16000	10	80 bytes	8,000 bytes/sec	BV32 (wideband) 32 kb/s
96-127	BV32/16000	20	120 bytes	6,000 bytes/sec	
96-127	BV32/16000	30	160 bytes	5,334 bytes/sec	
9	<none>	10	120 bytes	12,000 bytes/sec	G.722 at 64 kb/s using the Payload Type defined by IETF
9	<none>	20	200 bytes	10,000 bytes/sec	
9	<none>	30	280 bytes	9,334 bytes/sec	
96-127	G722-48/8000	10	100 bytes	10,000 bytes/sec	G.722 at 48 kb/s using dynamic payload type
96-127	G722-48/8000	20	160 bytes	8,000 bytes/sec	
96-127	G722-48/8000	30	220 bytes	7,333 bytes/sec	
96-127	G722-56/8000	10	110 bytes	11,000 bytes/sec	G.722 at 56 kb/s using dynamic payload type
96-127	G722-56/8000	20	180 bytes	9,000 bytes/sec	

Parameters from Session Description			Flowspec parameters		Comments
RTP/AVP code	rtpmap	ptime (msec)	Values b,m,M[1]	Values r,p[2]	
96-127	G722-56/8000	30	250 bytes	8,333 bytes/sec	
96-127	G722-64/8000	10	120 bytes	12,000 bytes/sec	G.722 at 64 kb/s using dynamic payload type
96-127	G722-64/8000	20	200 bytes	10,000 bytes/sec	
96-127	G722-64/8000	30	280 bytes	9,333 bytes/sec	
96-127	AMR-WB/16000	20	58 bytes	2,900 bytes/sec	AMR-WB at 6.6 kb/s
96-127	AMR-WB/16000	20	64 bytes	3,200 bytes/sec	AMR-WB at 8.85 kb/s
96-127	AMR-WB/16000	20	73 bytes	3,650 bytes/sec	AMR-WB at 12.65 kb/s
96-127	AMR-WB/16000	20	77 bytes	3,850 bytes/sec	AMR-WB at 14.25 kb/s
96-127	AMR-WB/16000	20	81 bytes	4,050 bytes/sec	AMR-WB at 15.85 kb/s
96-127	AMR-WB/16000	20	87 bytes	4,350 bytes/sec	AMR-WB at 18.25 kb/s
96-127	AMR-WB/16000	20	91 bytes	4,550 bytes/sec	AMR-WB at 19.85 kb/s
96-127	AMR-WB/16000	20	99 bytes	4,950 bytes/sec	AMR-WB at 23.05 kb/s
96-127	AMR-WB/16000	20	101 bytes	5,050 bytes/sec	AMR-WB at 23.85 kb/s
96-127	VMR-WB/16000[3]	20	42 bytes	2,100 bytes/sec	VMR-WB at 0.8kb/s
96-127	VMR-WB/16000[3]	20	43 bytes	2,150 bytes/sec	VMR-WB at 1.0 kb/s
96-127	VMR-WB/16000[3]	20	47 bytes	2,350 bytes/sec	VMR-WB at 2.7 kb/s
96-127	VMR-WB/16000[3]	20	50 bytes	2,500 bytes/sec	VMR-WB at 4.0 kb/s
96-127	VMR-WB/16000[3]	20	56 bytes	2,800 bytes/sec	VMR-WB at 6.2 kb/s
96-127	VMR-WB/16000[3]	20	62 bytes	3,100 bytes/sec	VMR-WB at 8.55 kb/s
96-127	VMR-WB/16000[3]	20	74 bytes	3,700 bytes/sec	VMR-WB at 13.3 kb/s

TABLE NOTES:
[1] b is bucket depth (bytes). m is minimum policed unit (bytes). M is maximum datagram size (bytes).
[2] r is bucket rate (bytes/sec). p is peak rate (bytes/sec).
[3] The Header-Free payload header format is assumed for the VMR-WB codec

7.6 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 specification, however, the codecs listed in this section 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 narrowband 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.

7.6.1 Supported Super-Wideband Codecs

The following sections describe super wideband codecs specified for use within PacketCable. 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 section MUST be met.

7.6.1.1 AAC

The Advanced Audio Codec (AAC) is specified by MPEG-2 Part 7 [ISO 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 14496-3]. Because of its use in the Apple iPod™ 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.[8] For this reason it is recommended for support within PacketCable.

AAC is a family of audio codecs sharing a common algorithmic base. The baseline AAC provides coding bit rates from 16 kb/s to 256 kb/s with sampling rates ranging from 7350Hz to 96kHz, though the most common rates are 24kHz, 32kHz, 44.1kHz, and 48kHz. 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) [TS 26.234], Multimedia Messaging Services (MMS) [TS 26.140], and Multimedia Broadcast/Multicast Services (MBMS) [TS 26.346]. 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.

7.6.1.1.1 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 kb/s (64 kb/s coding rate for each channel of stereo audio). This provides for audio quality comparable or better than MP3 at 160 kb/s. (The iTunes application provided by the Mac OS for creating AAC-LC files from CDs allows for user-selectable coding rates.)

AAC-LC incurs an algorithmic delay of around 100ms at rates >80 kb/s. Delay increases significantly at lower rates to over 300ms at 24 kb/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.

7.6.1.1.2 Enhanced AACPlus

The Enhanced AACPlus codec is specified by 3GPP in [TS 26.401], with the AAC encoder part specified in [TS 26.403]. Compared with AAC-LC, Enhanced AACPlus allows operation at a lower bit rate by incorporating Spectral Band Replication (SBR) [TS 26.404] and Parametric Stereo (PS) [TS 26.405] which are both defined in [ISO 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 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 section 8 [ISO 14496-3]. The result is a codec that provides for good quality stereo music transmission at rates as low as 24 kb/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 [TS 26.402]. 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 [ITUR BS 1534-1]. [TS 26.936] provides the results of this analysis and compares Enhanced AACPlus with the other 3GPP-mandated coder for audio applications (AMR-WB+). The 3GPP PSS [TS 26.234], MMS [TS 26.140], MBMS [TS 26.346] and IMS Messaging and Presence [TS 26.141] specifications also provide

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

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 [TS 26.402] for error concealment, stereo-to-mono down-mix, and re-sampling.

7.6.1.1.3 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 20ms. 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 kb/s to 128 kb/s per channel although can be as low as 16kb/s. Typical sampling rates used with AAC-LD are 32kHz to provide for a 14kHz frequency range and 48kHz for an audio bandwidth of 20kHz.

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 14496-3]. User Equipment and Media Servers supporting AAC-LD MUST allow a minimum coding rate during active speech of 48kb/s per audio channel with 64kb/s recommended.

7.6.1.1.4 Packet Loss Concealment

AAC includes decoder tools to conceal frame loss. For AAC, User Equipment and Media Servers MUST employ the error concealment tool.

7.6.1.1.5 Payload Header Format

For Packet-switched Streaming Service (PSS) [TS 26.234], 3GPP specifies the payload format described in [RFC 3016] for Enhanced AACPlus and other forms of AAC [RFC 3016]. 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) [TS 26.346], 3GPP specifies a more flexible payload format described in [RFC 3640]. ISMA [ISMA-2] and DVB [TS 102 005] also mandate the use of [RFC 3640] for the carriage of AAC over IP. Besides broader applicability, [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 [RFC 3016] and MAY support the payload format specified in [RFC 3640]. For any other service using AAC, User Equipment and Media Servers MUST support the payload format specified in [RFC 3640].

[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.

[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.

7.6.1.1.6 Session Description

[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 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 14496-3] section 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 14496-3] section 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.)

[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 14496-1] Table 9.
- The optional parameter "profile-level-id" defined by [ISO 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 14496-3] section 1.6.
- The parameters "sizeLength", "indexLength", and indexDeltaLength" must all be included and set as specified by [RFC 3640].

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

7.6.1.2 Extended AMR-WB (AMR-WB+)

Extended AMR-WB (AMR-WB+) [TS 26.290] 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) [TS 26.234]
- Multimedia Messaging Services (MMS) [[TS 26.140]
- Multimedia Broadcast/Multicast Services (MBMS) [TS 26.346]

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 [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 kHz to 19.2 kHz. AMR-WB+ provides coding bit rates of 5.2 to 36 kb/s for mono signals and 6.2 to 48 kb/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 [TS 26.190]. 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 [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 [ITUR BS 1534-1]. [TS 26.936] provides the results of this analysis and compares AMR-WB+ with the other 3GPP-mandated coder for audio applications (Enhanced AACPlus). The 3GPP PSS [TS 26.234], MMS [TS 26.140], MBMS [TS 26.346] and IMS Messaging and Presence [TS 26.411] specifications also provide guidance on the audio conditions under which AMR-WB+ provides higher performance than Enhanced AACPlus and vice versa.

7.6.1.2.1 Packet Loss Concealment

User Equipment and Media Servers supporting AMR-WB+ MUST use the methods defined in [TS 26.290] for frame erasure concealment.

7.6.1.2.2 Payload Header Format

User Equipment and Media Servers MUST support the payload header format as specified in [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.

7.6.1.2.3 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 [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

7.6.1.3 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 kb/s to 640 kb/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 (Center 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, Center, Right)
- 2-channel stereo with mono surround (Left, Right, Surround)
- 3-channel stereo with mono surround (Left, Center, Right, Surround)
- 4-channel quadraphonic (Left, Right, Left Surround, Right Surround)
- 5-channel surround (Left, Center, 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 Standards Committee) digital terrestrial television, DLNA (Digital Living Network Alliance) home networking. It is also an optional multi-channel 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 multi-channel Internet radio.

7.6.1.3.1 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.

7.6.1.3.2 Payload Header Format

User Equipment and Media Servers MUST support the payload header format for AC-3 specified in [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.

7.6.1.3.3 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 [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 that is equal to the audio sampling rate. Permitted rates are 32000 (32 kHz), 44100 (44.1 kHz), and 48000 (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 [RFC 3264].

7.6.1.4 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 kb/s to 6.144 Mbps, 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.

7.6.1.4.1 Packet Loss Concealment

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

7.6.1.4.2 Payload Header Format

User Equipment and Media Servers MUST support the payload header format for E-AC-3 specified in [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.

7.6.1.4.3 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 [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 [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'.

7.6.1.5 MP3

MP3 is a common reference to the MPEG-1 Part 3 Layer 3 [ISO 11172-3] and MPEG-2 Part 3 Layer 3 [ISO 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 48kHz. 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 320kbps.

7.6.1.5.1 Packet Loss Concealment

MP3 does not include any inherent packet loss concealment mechanism

7.6.1.5.2 Payload Header Format

User Equipment and Media Servers MUST support the payload header format for MP3 specified in [RFC 3555].

7.6.1.5.3 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 [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 create 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 sample rate and mode parameters are omitted from SDP, they are indicated in the payload.

7.6.2 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 in the rtpmap parameter:

Table 8 - 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 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.

7.7 Video Codec Specification

The PacketCable 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 PacketCable, and opens up new revenue streams for cable operators. From the perspective of cable customers, PacketCable 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.

PacketCable 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 PacketCable environment, this platform can provide increased video quality due to the higher bandwidth and enhanced QoS afforded by the PacketCable 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 handset, Wi-Fi IP phone and Wi-Fi-enabled PDA) 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 PacketCable, 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 TV display - especially with its high-definition capability and wide screen.

In addition, by standardizing video codecs, PacketCable 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 PacketCable 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, CIF, SD and HD.
2. Video Stream Direction: Different video applications deal with three different video stream directions: send only (1-way encode), receive only (1-way decode), and simultaneously send and receive (2-way encode/decode).

Since 1990, various video codecs have been developed to cater for different applications, mainly through two international standards bodies-VCEG (Video Coding Experts Group) of ITU-T and MPEG (Moving Picture Experts Group) of ISO/IEC. 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 H.262 as adopted by ITU-T) as the standard codec for DVD, digital TV and HDTV. Subsequent H-series and MPEG-series video codecs include H263 and MPEG-4 Part 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 bitstream, and a level places constraints on certain key parameters of the bitstream. 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, PacketCable 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 initiative, PacketCable video codec requirements need to be compatible with video codecs standardized by cellular standards bodies such as 3GPP and 3GPP2.

7.7.1 Supported Codecs

This section describes every video codec supported in PacketCable. 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 section MUST be met.

7.7.1.1 H.263

First approved in 1996, H.263 [H.263] is a video codec standardized by ITU-T VCEG 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 [H.261], the previous ITU-T standard for video telephony, in the areas of performance and error recovery. It has been designed to stream video at bandwidths as low as 20-24 Kbps. 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, as shown in [H.263 Annex X] it supports a wider range of video resolutions than H.261 (which only supports QCIF and CIF). As a result, H.263 has essentially replaced H.261.

7.7.1.1.1 Profile/Level Requirements

If User Equipment and Media Gateways support 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 PacketCable.

The above requirements are summarized in Table 9.

Table 9 - PacketCable 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 4 (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 [TS 26.235] and 3GPP packet-switched streaming services [TS 26.234], 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.

7.7.1.1.2 Payload Header Formats

The RTP payload format for H.263-encoded video media MUST be compliant with [RFC 4629].

7.7.1.1.3 Session Description

The session description for H.263-encoded video media MUST be compliant with [RFC 4629].

In particular, the MIME media type video/H263-2000 string is mapped to fields in the Session Description Protocol [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 bitstream 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 meaning are defined in [H.263 Annex X]. Note that the RTP payload format for H263- 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
```

7.7.1.2 H.264/AVC

H.264 [H.264] (or MPEG-4 Part 10), is a new generation of video codec jointly developed by the 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, 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., narrowband and wideband, wireline and wireless, broadcast, streaming, DVD storage, and video telephony). As reported in one benchmarking [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 4x4 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.

7.7.1.2.1 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 1.2 MUST be supported for CIF 1-way applications.
- H.264 Baseline Profile @ Level 1.2 MUST be supported for CIF 2-way applications.
- H.264 Baseline Profile @ Level 1.3 SHOULD be supported for CIF 1-way applications.
- H.264 Baseline Profile @ Level 1.3 SHOULD be supported for CIF 2-way applications.
- H.264 Main Profile @ Level 3 MUST be supported for SD 1-way applications.
- H.264 Baseline Profile @ Level 3 MUST be supported for SD 2-way applications.
- H.264 High Profile @ Level 4 MUST be supported for HD 1-way decode applications.

The H.264 support for HD 1-way encode and HD 2-way encode/decode is not specified for PacketCable.

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 10.

Table 10 - PacketCable 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 1.2, with constraint_set1_flag = 1 (MANDATORY) H.264 Baseline Profile @ Level 1.3, with constraint_set1_flag = 1 (RECOMMENDED)	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)	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, PacketCable 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 [TS 26.235] and H.264/AVC Baseline Profile Level 1.2, which is recommended for 3GPP packet-switched streaming services [TS 26.234]. The same requirements apply to 3GPP2.

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

7.7.1.2.2 Payload Header Formats

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

7.7.1.2.3 Session Description

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

In particular, the MIME media type video/H264 string is mapped to fields in the Session Description Protocol (SDP) [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
```

7.7.1.3 MPEG-2

MPEG-2 [ISO 13818-2] is a video codec standardized by ISO/IEC MPEG 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.

7.7.1.3.1 Profile/Level Requirements

If User Equipment or Media Gateways support MPEG-2, then the following requirements apply:

- MPEG-2 Main Profile @ Main Level SHOULD be supported for SD 1-way encode and decode applications.

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

The above requirements are summarized in Table 11.

Table 11 - PacketCable 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

7.7.1.3.2 Payload Header Formats

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

7.7.1.3.3 Session Description

The session description for MPEG-2-encoded video media MUST be compliant with [RFC 2250] and [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) [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) [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
```

7.7.1.4 MPEG-4 Part 2

MPEG-4 Part 2 [ISO 14496-2] is a video codec that belongs to the MPEG-4 standard family 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 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.

7.7.1.4.1 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 3 SHOULD be supported for CIF 1-way applications.
- MPEG-4 Part 2 Simple Profile @ Level 2 SHOULD be supported for CIF 2-way applications.

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

The above requirements are summarized in Table 12.

Table 12 - PacketCable 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 3 (RECOMMENDED)	MPEG-4 Part 2 Simple Profile @ Level 3 (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 recommendation, PacketCable 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 [TS 26.235] and MPEG-4 Part 2 Simple Profile Level 3, which is recommended for 3GPP packet-switched streaming services [TS 26.234]. The same requirements apply to 3GPP2.

7.7.1.4.2 Payload Header Formats

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

7.7.1.4.3 Session Description

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

In particular, the MIME media type video/MP4V-ES string is mapped to fields in the Session Description Protocol (SDP) [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 as 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 (90kHz), "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
```

7.7.2 Summary of Supported Codecs

Table 13 summarizes all video codecs that are supported by PacketCable, 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 13 - Summary of PacketCable 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 1.2, 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 3 (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 3 (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

7.7.3 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 specification.

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 specification does not specify any error-concealment and error-resilience algorithms which are not included in the mandated or recommended video codecs.

7.7.4 Codec Naming and FlowSpec Parameters for Video Codecs

The video codecs defined in this specification MUST be encoded with the string names in the rtpmap parameters as shown in Table 14.

Table 14 - 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
MPEG2 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 15 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 15 - Mapping of Video Codec Session Description Parameters to Flowspec

Parameters from Session Description			Flowspec parameters		Comments
RTP/AVP code	rtpmap	ptime (msec) (Note 1)	Values b,m,M[9] (Note 2)	Values r,p[10] (Note 3)	
H.263	H263 2000/90000	N/A	b, M = 1500 bytes m = 128 bytes	Level 45: 20K bytes/sec Level 40: 308K bytes/sec	b, M and m are default values. r, p = Max Compressed Bit Rate x 120%
H.264	H264/90000	N/A	b, M = 1500 bytes m = 128 bytes	Level 1b: 20K bytes/sec Level 1.2: 58K bytes/sec Level 1.3: 116K bytes/sec Level 3: 1.5M bytes/sec Level 4: 3M bytes/sec	b, M and m are default values. r, p = Max Compressed Bit Rate x 120%
MPEG-2	Transport Stream: MP2T/90000 Elementary Stream: MPV/90000	N/A	b, M = 1500 bytes m = 128 bytes	Main Level: 2.25M bytes/sec	b, M and m are default values. r, p = Max Compressed Bit Rate x 120%
MPEG-4 Part 2	MP4V-ES/90000	N/A	b, M = 1500 bytes m = 128 bytes	Level 0b: 9.6K bytes/sec Level 2: 19.2K bytes/sec	b, M and m are default values. r, p = Max Compressed Bit Rate x 120%

Notes:

(1) Ptime is not applicable to video in general. An application can set this parameter to an accurate value if it is known.

(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.

(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.

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

b= <modifier> : <bandwidth-value>

For example:

b= AS:99

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

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

[10] *r* is bucket rate (bytes/sec). *p* is peak rate (bytes/sec).

7.8 Media Quality Measurement and Monitoring

One of the principal goals of PacketCable is to enhance the user's experience with higher quality audio and video. Therefore, it is important for UEs to be able to monitor the quality of the audio and video streams. Depending on both the type of media and on who supplies the UE, the end user or the PacketCable network operator, it may be sufficient to monitor just the network performance or it may be beneficial to monitor the media quality in addition to the network quality. This section specifies the associated requirements to cover both cases.

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 specification focuses on objective metrics that can be obtained via non-intrusive testing.

7.8.1 RTCP XR VoIP Metrics Requirements

The RTCP XR VoIP Metrics [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 analyzers, or potentially by embedded monitoring functionality in CMTS and routers.

The RTCP XR VoIP Metrics include:

1. Network performance measurements that go beyond the basic loss and jitter measurements defined in the RTCP Receiver Report. The network performance measurements are useful in reporting round-trip delay, jitter buffer settings and, especially, quantifying packet loss,
2. Audio quality measurements that are useful when a PacketCable network operator is responsible for the provision of the media endpoints such as MTAs and MGs. The audio quality measurements include signal, noise and echo level measurements and objective speech quality measures.

The following table lists the types of metrics that fall under each measurement type. Audio quality measurements are currently only defined for narrowband audio.

Table 16 - Network Performance and Audio Quality Metrics as Defined in RFC 3611

Measurement Type	Reference	Metric Category	Metric
Network Performance Measurements	See Section 7.8.1.2	Packet Loss and Discard	Loss Rate Discard Rate Burst Loss Density Gap Loss Density Burst Duration Gap Duration Gmin
	See Section 7.8.1.3	Delay	Round Trip Delay End System Delay

Measurement Type	Reference	Metric Category	Metric
	See Section 7.8.1.6	Endpoint Configuration	PLC Type Jitter Buffer Type Jitter Buffer Rate Jitter Buffer - Nominal Delay Jitter Buffer - Maximum Delay Jitter Buffer - Absolute Max Delay
Audio Quality Measurements	See Section 7.8.1.4	Signal, Noise & Echo	Signal Level Noise Level Residual Echo Return Loss
	See Section 7.8.1.5	Call Quality	R Factor External R Factor MOS-LQ MOS-CQ

User Equipment and Media Gateways MAY support either RTCP XR Network performance measurements or RTCP XR Audio quality measurements or both. User Equipment and Media Gateways that support Network performance measurements MUST support all metrics defined to be within this measurement type by Table 16. User Equipment and Media Gateways that support Audio quality measurements MUST support all metrics defined to be within this measurement type by Table 16.

User Equipment and Media Gateways that support RTCP XR VoIP Metrics (of either measurement type) MUST exchange RTCP XR VoIP Metrics reports during active RTP sessions if negotiated. User Equipment and Media Gateways that support RTCP XR VoIP Metrics MUST concatenate RTCP XR payloads with RTCP SR and RR payloads, following rules for transmission intervals [RFC 3550] during active RTP sessions if negotiated.

An originating UE shall include the "rtcp-xr" attribute defined in [RFC 3611] in an SDP offer if all of the following conditions are true:

- The UE supports RTCP XR VoIP metrics,
- RTCP XR VoIP metrics are required for the session.

A terminating UE shall include the "rtcp-xr" attribute defined in [RFC 3611] in an SDP answer if all of the following conditions are true:

- The UE supports RTCP XR VoIP metrics,
- RTCP XR VoIP metrics are required for the session,
- The UE receives the "rtcp-xr" attribute in the related SDP offer.

The value of this attribute shall, at a minimum, indicate support of the RTCP XR VoIP metrics report block for all audio media streams. This support is indicated by including this attribute at the session level for a session consisting of audio streams only, or at each audio media level for sessions with diverse audio and non-audio media.

The direction of VoIP metrics reporting is the opposite of the direction of active media. For originating UEs sending sendrecv and recvonly offers, if the answer includes the "rtcp-xr" attribute with "voip-metrics" in its value field, the originating UE shall send VoIP metrics reports to the remote UE. For sendrecv and sendonly offers from an originating UE, if the answer includes the "rtcp-xr" attribute with "voip-metrics" in its value field, the originating UE shall expect VoIP metrics reports from the remote UE. For terminating UEs, if a sendrecv or recvonly answer includes the "rtcp-xr" attribute with "voip-metrics" in its value field, the terminating UE shall send VoIP metrics reports to the remote UE. If a sendrecv or sendonly answer from a terminating UE includes the "rtcp-xr" attribute with "voip-metrics" in its value field, the terminating UE expects VoIP metrics reports from the remote UE. VoIP metrics reports shall neither be sent nor expected if the answer does not include an "rtcp-xr" attribute with "voip-metrics" in its value field or if the answer indicates an inactive mode.

The frequency with which VoIP metrics reports may be sent by the UEs is limited by the RTCP bandwidth constraints as specified in [RFC 3556] for the RTP/AVP profile as defined in [RFC 3551]. Because of these constraints, it is possible that, for a short connection, no VoIP metrics blocks are sent in spite of SDP negotiation to the contrary.

Note that the SDP-based negotiation of the reporting of VoIP metrics via RTCP XR as defined in [RFC 3611] does not impact the minimal reporting of VoIP metrics via Sender Reports (SR) and Receiver Reports (RR) described in [RFC 3550]. These shall be provided by UEs, subject to RTCP bandwidth constraints defined in [RFC 3556].

User Equipment and Media Gateways that support the RTCP XR VoIP Metrics payload (of either measurement type) MUST measure or compute the reported values of the metrics as defined in [RFC 3611] and clarified in Sections 7.8.1.2 to 7.8.1.6 of this specification.

7.8.1.1 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 [ID SIP RTCP]. 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]. It is 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 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.

Regardless of the negotiation of RTCP-XR reporting via SDP, UEs shall PUBLISH locally computed VoIP metrics and metrics reported via Sender/Receiver Reports (SR/RR) to a collector as specified in [ID SIP RTCP]. The originating UE shall populate the request URI in the PUBLISH request with the Public User Identity of the originating user. The terminating UE shall populate the request URI in the PUBLISH request with the Public User Identity of the terminating user. In addition, if VoIP metrics reporting is enabled and if the UE receives RTCP XR VoIP metrics blocks from the remote UE, it shall PUBLISH the remote metrics as well. If VoIP metrics reporting is enabled then the UE shall PUBLISH local metrics, and remote metrics if available, at session termination or reconfiguration.

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 since session start or since 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 [RFC 3611].
- the local and remote versions of inter-arrival jitter, based on [RFC 3550].

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

7.8.1.2 Definition of Metrics related to Packet Loss and Discard

The VoIP Metrics [RFC 3611] payload contains six metrics related to packet or frame loss and discard. An average packet loss rate and an average packet discard rate report the proportion of packets lost or discarded on the call to date. A set of four burst parameters report the distribution of lost and discarded packets occurring during burst periods and gap periods.

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

Table 17 - Metrics Related to Packet Loss and Discard

METRIC	Description	Range
Loss Rate	Proportion of packets lost within the network	0 to 0.996
Discard Rate	Proportion of packets discarded due to late arrival	0 to 0.996
Burst Loss Density	Proportion of packets lost and discarded during burst periods	0 to 0.996
Gap Loss Density	Proportion of packets lost and discarded during gap periods	0 to 0.996
Burst Duration	Average length of burst periods (ms)	0 to 65,535
Gap Duration	Average length of gap periods (ms)	0 to 65,535
Gmin	Parameter used to define burst periods	0-255

User Equipment and Media Gateways when using RTCP XR MUST provide these parameters as defined in [RFC 3611].

7.8.1.3 Definition of Metrics Related to Delay

The VoIP Metrics payload includes two delay metrics [RFC 3611]. The Round Trip Delay is the delay between RTP interfaces, as typically measured using RTCP Sender Report (SR) or Receiver Report (RR) [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 18 - Metrics related to Delay

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

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

7.8.1.4 Definition of Metrics Related to Signal

The Signal Level, Noise Level and estimated Residual Echo Return Loss are intended to support the diagnosis of problems related to loss plan or PSTN echo. The intent is to report useful information that would typically be available from a vocoder or echo canceller rather than to impose the overhead of additional measurement algorithms on cost sensitive 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 this will in general not represent an accurate measurement of the residual echo but can provide a useful indication of the presence of echo problems. Echo occurring on the endpoint reporting this metric will be heard by the user at the remote endpoint, if significant delay is present on the call.

Table 19 - Metrics due to Signal

METRIC	Description	Range
Signal Level	RMS Signal level during active speech periods (dBm0) as defined in [P.56] and [P.561].	-30 to +3
Noise Level	RMS Noise level during silence periods (dBm0) as defined in [P.56] and [P.561].	-40 to -70
Residual Echo Return Loss	Estimated Echo Return Loss (after effects of echo canceller and NLP) from the local line echo canceller (dB) as defined in [G.168].	0 to 80

User Equipment and Media Gateways using RTCP XR with narrowband audio MUST provide Signal Level and Noise as defined in [G.107].

A PacketCable endpoint equipped with a line echo canceller and when using RTCP XR with narrowband audio MUST provide the Residual Echo Return Loss metric as defined in [RFC 3611].

7.8.1.5 Definition of Metrics related to Call Quality

Call quality metrics are useful when assessing the overall quality of a call [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 [G.107]) or in terms of Mean Opinion Score (MOS).

The maximum range of an R factor is 0-100 for narrowband voice transmission. Note, however, for wideband transmission the upper range can be greater than 100. The R factor defined in the ITU E Model is a conversational quality metric however it can be used to estimate conversational and listening quality MOS scores. The basic equation for determining an R Factor is:

R = Ro - Is - Id - I.e.,eff + A

Ro reflects the effects of noise and loudness, Is the effects of impairments occurring simultaneously with speech, Id the effects of delay related impairments and echo, I.e.,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. [G.107] defines an equation for converting an R factor into a MOS score; note however that this produces MOS scores slightly higher than those typically reported from subjective tests.

Table 20 - Metrics related to Call Quality

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

User Equipment and Media Gateways using RTCP-XR with narrowband audio 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 with narrowband audio MUST calculate R Factors using G.107 at a minimum [G.107].

User Equipment and Media Gateways using RTCP XR with narrowband audio 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 [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 [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 0.

Also the following equations below explain how to take measurements above and apply those to the E-model input parameters. For more information refer to [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 [G.107].

.User Equipment and Media Gateways using RTCP XR with narrowband audio MUST calculate the Ie,eff parameter using the function defined in [G.107]. However, User Equipment and Media Gateways MUST use the Ie and Bpl parameters defined in Table 21 for the codec and PLC combinations listed.

Table 21 - Ie and Bpl parameters for PacketCable Codecs

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

User Equipment and Media Gateways using RTCP XR with narrowband audio MUST calculate MOS-LQ using the R to MOS mapping function defined in [G.107] applied to the value (R - Id).

User Equipment and Media Gateways using RTCP XR with narrowband audio MUST calculate MOS-CQ using the R to MOS mapping function defined in [G.107] applied to the value R.

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

a. Use [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 P.862 to compare the impaired output files with the source files and average the results for each packet loss rate.
b. Determine the Ie value using the objective test scores for 0 percent loss. This may be obtained by iteratively searching for the Ie value that, when converted to an R factor and then an estimated P.862 score, gives the closest match to the measured P.862 score. Alternatively, the Ie value may be obtained by comparing the [P.862] score with other codecs with known Ie factor.
   $R_{adj} = R+(94-R)/3-3-115/(15+ABS(85-R))+40/(95-R)^2$
   Estimated PESQ score = $1 + 0.033.R_{adj} + R_{adj}(100-R_{adj})(R_{adj}-60)*0.000007$
c. Determine the Bpl value using the objective test scores for other packet loss rates. This may be obtained by iteratively searching for the Bpl value that, when converted to an R factor and then an estimated [P.862] score, gives the closest match to the measured [P.862] score. Alternatively, the Bpl value may be obtained by comparing the [P.862] score curve with other codecs with known Bpl factor.
d. It is generally advisable to compare the curve of estimated MOS score (derived per [G.107]) with available Absolute Category Rating (ACR) test data (if available) in order to verify values.

7.8.1.6 Definition of Parameters Related to Endpoint Configuration

The parameters in Table 22 describe some key configuration parameters of the PacketCable endpoint, that are useful in monitoring service quality and identifying some types of configuration related problems.

Table 22 - Parameters related to endpoint configuration

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

User Equipment and Media Gateways using RTCP XR MUST provide values to all parameters as defined in Table 22.

7.8.2 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 the document.

Annex A H.263 Profiles and Levels

The following tables summarize H.263 Profiles and Levels [H.263 Annex X].

Table 23 - 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 Mux*									
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 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*									
S: *Alternative Inter VLC*									

Annex/clause below for profile listed at right	0	1	2	3	4	5	6	7	8
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 24 - 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 bits/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 bits/s units	2	64	128	256

Annex B H.264/AVC Profiles and Levels

The following tables summarize H.264 Profiles and Levels [H.264]:

Table 25 - 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 26 - 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
8x8 vs. 4x4 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 Color Transform				X
Predictive Lossless Coding				X

Table 27 - 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 Kbps	4
1b	QCIF	15	128 Kbps	4
1.1	CIF or QCIF	7.5 (CIF)/30 (QCIF)	192 Kbps	2 (CIF)/9 (QCIF)
1.2	CIF	15	384 Kbps	6
1.3	CIF	30	768 Kbps	6
2	CIF	30	2 Mbps	6
2.1	HHR (480i or 576i)	30/25	4 Mbps	6

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
2.2	SD	15	4 Mbps	5
3	SD	30/25	10 Mbps	5
3.1	1280x720p	30	14 Mbps	5
3.2	1280x720p	60	20 Mbps	4
4	HD (720p or 1080i)	60p/30i	20 Mbps	4
4.1	HD (720p or 1080i)	60p/30i	50 Mbps	4
4.2	1920x1080p	60p	50 Mbps	4
5	2Kx1K	72	135 Mbps	5
5.1	2Kx1K or 4Kx2K	120/30	240 Mbps	5

Annex C Characteristics of Narrowband Codecs

The information provided in this annex is provided AS IS. CableLabs and the contributors to this annex make no representations or warranties, expressed or implied, including, but not limited to, non-infringement, accuracy, completeness, or fitness for any purpose. Through your use of the information contained in this annex you agree that CableLabs and the contributors to this annex shall not be liable for any cost, loss, damage (including, but not limited to, direct, indirect, incidental, consequential or punitive damages) or expense or costs of any kind arising from or related to the use of this annex.

Table 28 - Narrow Band Codecs (part 1)

Codec	Technology	Year Standar-dized	Coding Rate (kb/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 & 13.3	20 and 30	5 & 10	25 & 40	45 & 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	

Codec	Technology	Year Standar-dized	Coding Rate (kb/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)
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 29 - Narrowband Codecs (part 2)

Codec	Comp-lexity (MIPS) (Note 5)	Codec Impairment (G.107)	Calculated MOS CQE according to G.107 (Note 6)	MOS CQE for intra-MSO calls (Note 7)	MOS CQE for inter-MSO calls (Note 8)	MOS CQE for Cable to PSTN calls (Note 9)	MOS CQE for Cable to Cellular calls (Note 9)	Known comparison with references in official 3rd party listening MOS tests (Note 10)	Packet Loss Rate for 0.5 MOS degradation (Note 11)
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%

Codec	Comp-lexity (MIPS) (Note 5)	Codec Impairment (G.107)	Calculated MOS CQE according to G.107 (Note 6)	MOS CQE for intra-MSO calls (Note 7)	MOS CQE for inter-MSO calls (Note 8)	MOS CQE for Cable to PSTN calls (Note 9)	MOS CQE for Cable to Cellular calls (Note 9)	Known comparison with references in official 3rd party listening MOS tests (Note 10)	Packet Loss Rate for 0.5 MOS degradation (Note 11)
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 kb/s)	4.29 (at 12.2kb/s)	4.29 (at 12.2kb/s)	4.24 (at 12.2kb/s)	4.29 (at 12.2kb/s)	4.23 (at 12.2kb/s)	~ G.726(32K)	3%
IS-893 SMV (Note 12)	~40 WMOPS							=G.711 (with Mode_0) >EVRC (with Mode_0) ~G.711 (with Mode_1)	3%

Table 30 - Narrowband Codecs (part 3)

Codec	Performance for Background Noise	MOS Reduction for Tandem Encodings (Note 13)	Performance	Other Functionality	Preval-ence	Applications	Status	Reference Fixed Point C Code
G.711	Toll Quality	Toll Quality	Very good for speech, audio, DTMF, text, fax & voiceband data.	PLC defined in App. I, SID defined in App. II	High	ubiquitous	ITU-T	Algorithmic Description available
G.726	Toll Quality	Toll Quality	Good for speech, audio & DTMF; very good for text; poor for fax & modem.		High	International links, DCME	ITU-T	Algorithmic Description available
G.728	Toll Quality	Toll Quality	Good for speech		Medium	DCME, video conferencing, PacketCable E-MTA	ITU-T	Algorithmic Description available
G.729	<=Toll Quality	< Toll Quality	Good for speech without transcoding	Integrated VAD and PLC	High	IP phones	ITU-T	Free fixed Point C-Code
G.729A	<= Toll Quality	< Toll Quality	Good for speech without transcoding	Integrated VAD and PLC	High	IP phones, DSVD	ITU-T	Free fixed Point C-Code

Codec	Performance for Background Noise	MOS Reduction for Tandem Encodings (Note 13)	Performance	Other Functionality	Preval-ence	Applications	Status	Reference Fixed Point C Code
G.729E	Toll Quality	Toll Quality	Good for speech	Integrated VAD and PLC; Music Detection	Low	PacketCable E-MTA	ITU-T	Free fixed Point C-Code e
G.723.1	<= Toll Quality	< Toll Quality	Acceptable for speech at 6.3kb/s; poor at 5.3kb/s		Medium	videophone over dial-up, IP phones	ITU-T but used in enterprise networks	Free fixed Point C-Code
iLBC	Toll Quality	< Toll Quality	Good for speech without transcoding	Integrated VAD and PLC	Medium	PacketCable E-MTA, Skype	IETF; PacketCable E-MTA	Fixed Point C-Code available free for PacketCable IPR vendors.
BV16	Toll Quality	Toll Quality	Good for speech	Example PLC available	Low	PacketCable E-MTA	PacketCable E-MTA	Algorithmic Description available; Fixed Point C-Code available for licensing
GSM EFR	Toll Quality	Toll Quality	Good for speech	Proprietary VQE - Voice Quality Enhancement.	High	Cellular	GSM and 3GPP	Free fixed Point C-Code
TDMA IS-641	< Toll Quality	< Toll Quality	Good for speech without transcoding		Medium	Cellular	TDMA	Free fixed Point C-Code
EVRC IS-127	< Toll Quality	< Toll Quality	Good for speech without transcoding		High	Cellular	CDMA/3GPP2	Free fixed Point C-Code
IS-733	Toll Quality	Toll Quality	Good for speech without transcoding		High	Cellular	CDMA/3GPP2	Free fixed Point C-Code
AMR	Toll Quality (at 12.2kb/s)	Toll Quality (at 12.2kb/s)	Good for speech (at 12.2kb/s)	Proprietary VQE	High	Cellular	GSM and 3GPP	Free fixed Point C-Code
IS-893 SMV (Note14)	Toll Quality built-in noise suppression	Toll Quality (Mode_0)	Good for speech	Noise Suppression	Low	Cellular	3GPP2	Free fixed Point C-Code e

Narrowband Codecs (part 1-3) Table Notes

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, while Mode_1 be as good as EVRC.

Note 5: Complexity is reported as MIPS (Million Instructions Per Second) and stated computational complexity numbers include one encoder and one decoder for the TI TMS320C54x architecture; SMV reports as WMOPS

Note 6: This MOS CQE score is calculated based on Ie values according to ITU specification 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 7: 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.

Note 8: 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 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 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 10: Per ITU-T listening MOS procedure stipulated in P.800, 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 scores are pretty far away (like 0.4 or more).
- < : denotes that MOS scores are meaningfully worse, like 0.2 typical.
- ~ : denotes that MOS scores are somewhat less, however, the differences are not statistically meaningful (typically within 0.1).
- = : denotes that MOS scores are really comparable, could actually be higher than the reference codec (but not statistically meaningful).
- > : denotes that MOS scores are better than reference codec.

Note 11: 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 12: SMV = Selectable Mode Vocoder. Design goals set forth by 3GPP2/TIA are Mode_0 to be better than EVRC under all conditions, while Mode_1 be as good as EVRC.

Note 13: Tandem encodings means two encodings using the same codec with G.711 in between.

Note 14: SMV = Selectable Mode Vocoder. Design goals set forth by 3GPP2/TIA are Mode_0 to be better than EVRC under all conditions, while Mode_1 be as good as EVRC.

Annex D Characteristics of Wideband Codecs

The information provided in this annex is provided AS IS. CableLabs and the contributors to this annex make no representations or warranties, expressed or implied, including, but not limited to, non-infringement, accuracy, completeness, or fitness for any purpose. Through your use of the information contained in this annex you agree that CableLabs and the contributors to this annex shall not be liable for any cost, loss, damage (including, but not limited to, direct, indirect, incidental, consequential or punitive damages) or expense or costs of any kind arising from or related to the use of this annex.

Table 31 - Wideband Codecs (part 1)

Codec	Technology	Speech Model Used	Year Standar-dized	Sampling Rate	Audio Bandwidth	Coding Rate (kb/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 to 7kHz	48, 56, 64	No	0.0625	0
G.722.1	MLT (Modulated Lapped Transform)	No	1999	16 kHz	50 to 7kHz	24, 32	No	20	20
G.722.2 / AMR-WB	ACELP	YES	2001/2002	16 kHz	50 to 6.4kHz(Note 1)	6.60, 8.85, 12.65, 14.25, 15.85, 18.25, 19.85, 23.05, 23.85	Defined in 3GPP TS 26.093	20	5
VMR-WB	ACELP	YES	2004	16 kHz	60 to 6.4kHz (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 to 7kHz	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	50Hz to >7 kHz (Note 8)	Avg 80	Yes	10,20,30,40	0
iSAC (GIPS)	Transform Coding	No	not an approved standard	16 kHz	50Hz to >7kHz (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	20Hz to >7kHz (Note 8)	32	No	5	0

Table 32 - Wideband Codecs (part 2)

Codec	Encoder & 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)	Comp-lexity	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 kbps: okay for audio 48 kbps: marginal quality (noisy)
G.722.1	0	40	60	5.5	15.7	10.3 WMOPS	24 kbps < G.722 @ 56 kbps; 24 kbps ≥ G.722 @ 48 kbps; 32 kbps < G.722 @ 64 kbps; (32 kbps ≥ G.722 @ 56 kbps except tandeming and -16 dBov)	1.5% for G.722.1 @ 24 kbps 1.0% for G.722.1 @ 32 kbps	Relatively good for music
G.722.2 / AMR-WB	1.875	26.875	46.875	5.3	23	38 WMOPS	6.6, 8.85 kbps < G.722 @ 48 15.85 kbps ≥ G.722 @ 56 23.85 kbps ≥ G.722 @ 64	2.0% for 12.65 & 15.85 kbps 1.8% for 19.85 & 23.85 kbps	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

Codec	Encoder & 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)	Comp-lexity	Output Quality versus G.722 for Clean Speech (Note 7)	Packet Loss Rate for 0.5 MOS Degradation	Performance for Music
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 kbps	5.3%	Okay for audio

Table 33 - Wideband Codecs (part 3)

Codec	Other Functionality	Prevalence in Wideband-Capable Systems *	Integrated Packet Loss Concealment	Integrated VAD/DTX/CNG	Status	Reference Fixed Point C Code
G.722		High	NO	NO	ITU-T	Fixed Point C-Code available for free
G.722.1		Low	Frame Repeat specified in G.722.1	No	ITU-T	Fixed Point C-Code available for free
G.722.2 / AMR-WB		Mandatory 3GPP WB codec soon to be deployed	YES	YES	Same as 3GPP TS. 26.190	Fixed Point C-Code available for free
VMR-WB	Integrated Noise suppression	Recently standardized by 3GPP2	YES	YES	3GPP2 C.R0052-0	Fixed Point C-Code available for free
SMV-WB	Integrated Noise suppression	Low	YES	YES	SMV-NB is in 3GPP2 IS-893 SMV-WB is similar technology	Fixed Point C-Code available for free
iPCM-wb (GIPS)	Noise cancellation / suppression available with VQE/NetEQ	Low	optional VQE/NetEQ module	optional with VQE/NetEQ module	proprietary	Fixed Point C-Code available for licensing
iSAC (GIPS)	Noise cancellation / suppression available with VQE/NetEQ	Very high (Skype, Google, AOL, Yahoo, QQ…)	optional VQE/NetEQ module	optional with VQE/NetEQ module	proprietary	Fixed Point C-Code available for licensing

Codec	Other Functionality	Prevalence in Wideband-Capable Systems *	Integrated Packet Loss Concealment	Integrated VAD/DTX/CNG	Status	Reference Fixed Point C Code
BV32 (Broadcom)		Low	YES	NO	BV16 is in PacketCable E-MTA; BV32 is Similar technology	Fixed Point C-Code available for licensing

Note 1 G.722.2/AMR-WB only encodes 0 - 6.4 kHz band; the 6.4 - 7 kHz band is estimated and synthesized based on low-band information, except the 23.85 kbps 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 based on 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 8kHZ

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 kb/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+.

Appendix I Acknowledgements

This Specification was developed and influenced by numerous individuals representing many different vendors and organizations. CableLabs hereby wishes to thank everybody who participated directly or indirectly in this effort.

CableLabs wishes to recognize the following individuals for their significant involvement and contributions to the issued specification:

Hisham Abdelhamid - Cisco
John Atkinson - Nortel
Juin-Hwey (Raymond) Chen - Broadcom
Harprit S. Chhatwal - Nuera
Keith Chu - Mindspeed
Chad Griffiths - Broadcom
Rajesh Kumar - Cisco
Satish Kumar - Texas Instruments
Keith Lantz - Cisco
Gordon Li - Broadcom
Joe Stone - Cisco
Venkatesh Sunkad - CableLabs
Derek Underwood - lead editor - Nokia Siemens Networks

Steve Dotson- CableLabs

Appendix II Revision History

The following Engineering Change Notices were incorporated in PKT-SP-CODEC-MEDIA-I02-061013.

ECN	ECN Date	Summary
CODEC-MEDIA-N-06.0333-4	9/18/06	Clarifications to Quality Metrics, Minor Editorial Changes
CODEC-MEDIA-N-06.0347-4	9/18/06	Removal of mptime, Addition of maxptime, Packetization Rate Clarifications, Minor Technical and Editorial Changes
CODEC-MEDIA-N-06.0376-3	10/9/06	Editorial changes to remove optional information related to codecs
CODEC-MEDIA-N-06.0377-1	10/12/06	EC to correct ECN 06.0347-4.

The following Engineering Change Notices were incorporated in PKT-SP-CODEC-MEDIA-I03-070925.

ECN	ECN Date	Summary
CODEC-MEDIA-N-07.0412-3	5/7/07	Addition of super-wideband codecs
CODEC-MEDIA-N-07.0443-2	6/25/07	Updating of references and alignment with Release 7 of 3GPP IMS specifications
CODEC-MEDIA-N-07.0455-3	6/25/07	Specification cleanup - minor technical and editorial updates
CODEC-MEDIA-N-07.0483-1	8/20/07	Updating of references for R7 alignment of Codec and Media specification