PacketCable™ 1.5 Specifications

Audio/Video Codecs

PKT-SP-CODEC1.5-I01-050128

ISSUED

Notice

This PacketCable specification is a cooperative effort undertaken at the direction of Cable Television Laboratories, Inc. (CableLabs®) for the benefit of the cable industry. Neither CableLabs, nor any other entity participating in the creation of this document, is responsible for any liability of any nature whatsoever resulting from or arising out of use or reliance upon this document by any party. This document is furnished on an AS-IS basis and neither CableLabs, nor other participating entity, provides any representation or warranty, express or implied, regarding its accuracy, completeness, or fitness for a particular purpose.

Document Status Sheet

Document Control Number:	PKT-SP-CODEC1.5-I01-050128			
Document Title:	Audio/Video Codecs			
Revision History:	D01 – September 30, 2004 D02 – December 10, 2004 I01 – January 28, 2005			
Date:	January 28, 2005			
Status:	~~Work in Progress~~	~~Draft~~	Issued	~~Closed~~
Distribution Restrictions:	~~Author 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

List of Tables

This page intentionally left blank.

1 SCOPE

This document addresses interfaces between PacketCable™ client devices for audio and video communication. Specifically, it identifies the audio and video codecs necessary to provide the highest quality and the most resource-efficient service delivery to the customer. This document also specifies the performance required in client devices to support future PacketCable codecs. Additionally, this document describes a suggested methodology for optimal network support for codecs.

1.1 Introduction and Overview

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

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

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

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

1.2 Purpose of the Document

The purpose of this document is to extend the existing PacketCable 1.0 Codec specification by introducing two new low-bit codecs, ITU-T T.38 fax relay for reliable fax transmission, IETF RFC 2833 DTMF Relay for reliable DTMF transmission, and metrics to measure voice quality. It is issued to facilitate design and field-testing leading to the manufacturability and interoperability of conforming hardware and software by multiple vendors.

1.3 Organization of Document

This document covers the following major topics:

- Network issues affecting and influenced by codecs, along with a discussion of codec implications on network design (Section 5).
- Client device requirements necessary to support codecs (Section 6).
- Audio codec specifications (Section 7).
- Video codec requirements and specifications (Section 8).
- RTP and RTCP usage (Section 9).

1.4 Requirements Syntax

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

"MUST"	This word or the adjective "REQUIRED" means that the item is an absolute requirement of this specification.
"MUST NOT"	This phrase means that the item is an absolute prohibition of this specification.
"SHOULD"	This word or the adjective "RECOMMENDED" 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 or the adjective "OPTIONAL" means that this item is truly optional. One vendor may choose to include the item because a particular marketplace requires it or because it enhances the product, for example; another vendor may omit the same item.

Other text is descriptive or explanatory.

The legal/regulatory classification of IP-based voice communications provided over cable networks and otherwise, and the legal/regulatory obligations, if any, borne by providers of such voice communications, are not yet fully defined by appropriate legal and regulatory authorities. Nothing in this specification is addressed to, or intended to affect, those issues. In particular, while this document uses standard terms such as "call," "call signaling," "telephony," etc., it will be evident from this document that while a PacketCable network performs activities analogous to these PSTN functions, the manner by which it does so differs considerably from the manner in which they are performed in the PSTN by telecommunications carriers. These differences may be significant for legal/regulatory purposes.

1.5 Phasing of Requirements

The codec requirements contained in this specification cover both audio and video Multimedia Terminals (MTAs) and Trunking Gateways (Media Gateway). The term MTA-2 is used to define a terminal supporting video.

In the initial phase of PacketCable, MTAs are not required to support the MTA-2 requirements as defined in Section 8 of this specification. MTAs MUST support the requirements for audio terminals as defined in Sections 5, 6, 7, and 9.

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

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. Notwithstanding, intellectual property rights may be required to use or implement such normative references.

[1] PacketCable 1.5 Architecture Framework Technical Report, PKT-TR-ARCH1.5 V01-050128, January 28, 2005.

[2] PacketCable 1.5 Dynamic Quality of Service Specification, PKT-SP-DQOS1.5 I01-050128, January 28, 2005.

[3] PacketCable 1.5 Network-Based Call Signaling Protocol Specification, PKT-SP-NCS1.5 I01-050128, January 28, 2005.

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

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

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

[7] IETF RFC 1890, RTP Profile for Audio and Video Conferences with Minimal Control, January 1996.

[8] IETF RFC 2327, SDP: Session Description Protocol, April 1998.

[9] IETF RFC 0821, Simple Mail Transfer Protocol, August 1982.

[10] IETF RFC 2104, HMAC: Keyed-Hashing for Message Authentication, February 1997.

[11] IETF RFC 1945, Hypertext Transfer Protocol -- HTTP/1.0, May 1996.

[12] IETF RFC 2068, Hypertext Transfer Protocol -- HTTP/1.1, January 1997.

[13] IETF RFC 3435, Media Gateway Control Protocol (MGCP) Version 1.0, January 2003.

[14] IETF RFC 2138, Remote Authentication Dial In User Service (RADIUS), April 1997.

[15] IETF RFC2139, RADIUS Accounting,April 1997.

[16] IETF RFC 1889, RTP: A Transport Protocol for Real-Time Applications, January 1996.

[17] ITU-T Rec. G.711, Pulse Code Modulation (PCM) Of Voice Frequencies, November 1988.

[18] ITU-T Rec. G.165, Echo Cancellers, March 1993.

[19] ITU-T Rec. G.168, Digital Network Echo Cancellers, June 2002.

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

[21] ITU-T Rec. V.18, Operational and Interworking Requirements for DCEs Operating in the Text Telephone Mode, November 2000.

[22] ITU-T Rec. G.728, Coding of Speech at 16 kbit/S Using Low-Delay Code Excited Linear Prediction, September 1992.

[23] ITU-T Rec. G.729 Annex E, Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear-prediction (CS-ACELP) 11.8 kbit/s CS-ACELP speech coding algorithm, September 1998.

[24] IETF RFC 2833, RTP Payload for DTMF Digits, Telephony Tones and Signals, May 2000.

[25] Telcordia, "LSSGR: Signaling for Analog Interfaces", Generic Requirements Gr–506–Core Issue 1 June 1996, Revision 1 November 1996, Morristown, New Jersey.

[26] SIA DC-05-1999.09, Digital Communication Standard - Ademco ® Contact ID Protocol - for Alarm System Communications", Security Industry Association Copyright 1999 - Ademco Group Publication Order Number: 14085.

[27] ETSI ES 201 235-2 V1.2.1 (2002-05) ETSI Standard "Access and Terminals (AT); Specification of Dual-Tone Multi-Frequency (DTMF) Transmitters and Receivers; Part 2: Transmitters."

[28] ETSI ES 201 235-3 V1.2.1 (2002-05) ETSI Standard "Access and Terminals (AT); Specification of Dual-Tone Multi-Frequency (DTMF) Transmitters and Receivers; Part 3: Receivers."

[29] IETF RFC 3611 Real Time Control Protocol Extended Reports (RTCP XR), T. Friedman et al., November 2003.

[30] ITU Rec. P.56 – Objective Measurement of Active Speech Level, 1993.

[31] ITU Rec. P.561 – In-Service, Non-Intrusive Measurement Device, 1996.

[32] ITU Rec. G.107 - The E Model, a computational model for use in transmission planning, May 2003.

[33] ITU 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.

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

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

[36] iLBC Speech Codec Fixed Point Reference Code for PacketCable, Version 1.0.3, October 2003. (For further information, contact PacketCable, Cable Television Laboratories, Inc.)

[37] IETF draft "RTP Payload Format for BroadVoice Speech:, draft- ietf-avt-bv-02.txt, June 2005, J. Chen, et al.

2.2 Informative References

[38] PacketCable 1.5 Security Specification, PKT-SP-SEC1.5 I01-050128, January 28, 2005.

[39] PacketCable 1.5 Event Messaging Specification, PKT-SP-EM1.5 I01-050128, January 28, 2005.

[40] PacketCable 1.5 MTA Device Provisioning Specification, PKT-SP-PROV1.5 I01-050128, January 28, 2005.

[41] PacketCable 1.5 PSTN Gateway Call Signaling Protocol Specification, PKT-SP-TGCP1.5 I01-050128, January 28, 2005.

[42] PacketCable 1.5 Audio Server Protocol Specification, PKT-SP-ASP1.5-I01-050128, January 28, 2005.

[43] BroadVoice™16 Speech Codec Specification, Version 1.2, October 2003, J.Chen et al. (For further information, contact PacketCable, Cable Television Laboratories, Inc.).

[44] ITU-T Rec. G.729 Annex A, Reduced Complexity 8 kbit/s CS-ACELP Speech Codec, November 1996.

[45] 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.

[46] ITU-T Rec. H.245, Control protocol for multimedia communication, July 2003.

[47] ITU-T Rec. H.261, Video codec for audiovisual services at p x 64 kbit/s, March 1993.

[48] ITU-T Rec. H.263, Video coding for low bit rate communication, February 1998.

[49] ITU-T Rec. H.323, Packet-based multimedia communications systems, July 2003.

[50] ITU-T Rec. H.324, Terminal for low bit-rate multimedia communication, March 2002.

2.3 Reference Acquisition

- ATIS Document center: www.atis.org
- ETSI Standards: www.etsi.org/services_products/e-shop/home.htm
- IETF RFCs: www.ietf.org/
- ITU-T Recommendations: www.itu.int/ITU-T/publications/recs.html
- PacketCable specifications: Cable Television Laboratories, www.packetcable.com/specifications/
- Telcordia: http://www.telcordia.com/services/testing/lab_access/gr-listing.html
- Security Industry Association: www.siaonline.org

3 TERMS AND DEFINITIONS

This specification uses the following terms:

Access Control	Limiting the flow of information from the resources of a system only to authorized persons, programs, processes or other system resources on a network.
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.
Admitted	A service flow is said to be "admitted" when the CMTS has reserved resources (e.g., bandwidth) for it on the DOCSIS network.
A-link	A-Links are SS7 links that interconnect STPs and either SSPs or SCPs. A' stands for "Access."
Asymmetric Key	An encryption key or a decryption key used in public key cryptography, where encryption and decryption keys are always distinct.
Audio Server	An Audio Server plays informational announcements in PacketCable network. Media announcements are needed for communications that do not complete and to provide enhanced information services to the user. The component parts of Audio Server services are Media Players and Media Player Controllers.
Authentication	The process of verifying the claimed identity of an entity to another entity.
Authenticity	The ability to ensure that the given information is without modification or forgery and was in fact produced by the entity that claims to have given the information.
Authorization	The act of giving access to a service or device if one has the permission to have the access.
Cipher	An algorithm that transforms data between plaintext and ciphertext.
Ciphersuite	A set, which must contain both an encryption algorithm and a message authentication algorithm (e.g., a MAC or an HMAC). In general, it may also contain a key management algorithm, which does not apply in the context of PacketCable.
Ciphertext	The (encrypted) message output from a cryptographic algorithm that is in a format that is unintelligible.
Cleartext	The original (unencrypted) state of a message or data. Also called plaintext.
Confidentiality	A way to ensure that information is not disclosed to any one other then the intended parties. Information is encrypted to provide confidentiality. Also known as privacy.
Cryptanalysis	The process of recovering the plaintext of a message or the encryption key without access to the key.
Cryptographic algorithm	An algorithm used to transfer text between plaintext and ciphertext.
Decipherment	A procedure applied to ciphertext to translate it into plaintext.
Decryption	A procedure applied to ciphertext to translate it into plaintext.
Decryption key	The key in the cryptographic algorithm to translate the ciphertext to plaintext.
Digital certificate	A binding between an entity's public key and one or more attributes relating to its identity, also known as a public key certificate.

Digital signature	A data value generated by a public key algorithm based on the contents of a block of data and a private key, yielding an individualized cryptographic checksum.
Downstream	The direction from the head-end toward the subscriber location.
Encipherment	A method used to translate information in plaintext into ciphertext.
Encryption	A method used to translate information in plaintext into ciphertext.
Encryption Key	The key used in a cryptographic algorithm to translate the plaintext to ciphertext.
Endpoint	A Terminal, Gateway or MCU.
Errored Second	Any 1-sec interval containing at least one bit error.
Event Message	Message capturing a single portion of a connection.
F-link	F-Links are SS7 links that directly connect two SS7 end points, such as two SSPs. 'F' stands for "Fully Associated."
Flow [DOCSIS Flow]	(a.k.a. DOCSIS-QoS "service flow") A unidirectional sequence of packets associated with a SID and a QoS. Multiple multimedia streams may be carried in a single DOCSIS Flow.
Flow [IP Flow]	A unidirectional sequence of packets identified by ISO Layer 3 and Layer 4 header information. This information includes source/destination IP addresses, source/destination port numbers, protocol ID. Multiple multimedia streams may be carried in a single IP Flow.
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.
Integrity	A way to ensure that information is not modified except by those who are authorized to do so.
IntraLATA	Within a Local and Access Transport Area.
Jitter	Variability in the delay of a stream of incoming packets making up a flow such as a voice communication.
Kerberos	A secret-key network authentication protocol that uses a choice of cryptographic algorithms for encryption and a centralized key database for authentication.
Key	A mathematical value input into the selected cryptographic algorithm.
Key Exchange	The swapping of public keys between entities to be used to encrypt communication between the entities.
Key Management	The process of distributing shared symmetric keys needed to run a security protocol.
Key Pair	An associated public and private key where the correspondence between the two are mathematically related, but it is computationally infeasible to derive the private key from the public key.

Keying Material	A set of cryptographic keys and their associated parameters, normally associated with a particular run of a security protocol.
Keyspace	The range of all possible values of the key for a particular cryptographic algorithm.
Latency	The time, expressed in quantity of symbols, taken for a signal element to pass through a device.
Link Encryption	Cryptography applied to data as it travels on data links between the network devices.
Network Layer	Layer 3 in the Open System Interconnection (OSI) architecture that provides network information that is independent from the lower layers.
Network Management	The functions related to the management of data across the network.
Network Management OSS	The functions related to the management of data link layer and physical layer resources and their stations across the data network supported by the hybrid fiber/coax system.
Nonce	A random value used only once that is sent in a communications protocol exchange to prevent replay attacks.
Non-Repudiation	The ability to prevent a sender from denying later that he or she sent a message or performed an action.
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.
Plaintext	The original (unencrypted) state of a message or data. Also called cleartext.
Pre-shared Key	A shared secret key passed to both parties in a communication flow, using an unspecified manual or out-of-band mechanism.
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.
Private Key	The key used in public key cryptography that belongs to an individual entity and must be kept secret.
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.
Public Key	The key used in public key cryptography that belongs to an individual entity and is distributed publicly. Other entities use this key to encrypt data to be sent to the owner of the key.
Public Key Certificate	A binding between an entity's public key and one or more attributes relating to its identity, also known as a digital certificate.
Public Key Cryptography	A procedure that uses a pair of keys, a public key and a private key for encryption and decryption, also known as an asymmetric algorithm. A user's public key is publicly available for others to use to send a message to the owner of the key. A user's private key is kept secret and is the only key that can decrypt messages sent encrypted by the user's public key.

Root Private Key	The private signing key of the highest-level Certification Authority. It is normally used to sign public key certificates for lower-level Certification Authorities or other entities.
Root Public Key	The public key of the highest level Certification Authority, normally used to verify digital signatures generated with the corresponding root private key.
Secret Key	The cryptographic key used in a symmetric key algorithm, which results in the secrecy of the encrypted data depending solely upon keeping the key a secret, also known as a symmetric key.
Session Key	A cryptographic key intended to encrypt data for a limited period of time, typically between a pair of entities.
Signed and Sealed	An "envelope" of information which has been signed with a digital signature and sealed using encryption.
Subflow	A unidirectional flow of IP packets characterized by a single source and destination IP address and source and destination UDP/TCP port.
Symmetric Key	The cryptographic key used in a symmetric key algorithm, which results in the secrecy of the encrypted data depending solely upon keeping the key a secret, also known as a secret key.
Systems Management	Functions in the application layer related to the management of various open systems Interconnection (OSI) resources and their status across all layers of the OSI architecture.
Transit Delays	The time difference between the instant at which the first bit of a PDU crosses one designated boundary, and the instant at which the last bit of the same PDU crosses a second designated boundary.
Trunk	An analog or digital connection from a circuit switch that carries user media content and may carry voice signaling (M_F, R_2, etc.).
Tunnel Mode	An IPSec (ESP or AH) mode that is applied to an IP tunnel, where an outer IP packet header (of an intermediate destination) is added on top of the original, inner IP header. In this case, the ESP or AH transform treats the inner IP header as if it were part of the packet payload. When the packet reaches the intermediate destination, the tunnel terminates and both the outer IP packet header and the IPSec ESP or AH transform are taken out.
Upstream	The direction from the subscriber location toward the head-end.
X.509 certificate	A public key certificate specification developed as part of the ITU-T X.500 standards directory.

4 ABBREVIATIONS AND ACRONYMS

This specification uses the following abbreviations:

AAA	Authentication, Authorization and Accounting
AES	Advanced Encryption Standard. A block cipher, used to encrypt the media traffic in PacketCable.
AF	Assured Forwarding. This is a DiffServ Per Hop Behavior.
AH	Authentication header. An IPSec security protocol that provides message integrity for complete IP packets, including the IP header.
AMA	Automated Message Accounting. A standard form of call detail records (CDRs) developed and administered by Bellcore (now Telcordia Technologies).
ASD	Application-Specific Data. A field in some Kerberos key management messages that carries information specific to the security protocol for which the keys are being negotiated.
AT	Access Tandem
ATM	Asynchronous Transfer Mode. A protocol for the transmission of a variety of digital signals using uniform 53-byte cells.
BAF	Bellcore AMA Format, also known as AMA.
BCID	Billing Correlation ID
BPI+	Baseline Privacy Plus Interface Specification. The security portion of the DOCSIS 1.1 standard that runs on the MAC layer.
CA	Certification Authority. A trusted organization that accepts certificate applications from entities, authenticates applications, issues certificates and maintains status information about certificates.
CA	Call Agent. The part of the CMS that maintains the communication state, and controls the line side of the communication.
CBC	Cipher Block Chaining Bode. An option in block ciphers that combine (XOR) the previous block of ciphertext with the current block of plaintext before encrypting that block of the message.
CBR	Constant Bit Rate
CDR	Call Detail Record. A single CDR is generated at the end of each billable activity. A single billable activity may also generate multiple CDRs.
CIC	Circuit Identification Code. In ANSI SS7, a two-octet number that uniquely identifies a DSO circuit within the scope of a single SS7 Point Code.
CID	Circuit ID (Pronounced "kid"). This uniquely identifies an ISUP DS0 circuit on a Media Gateway. It is a combination of the circuit's SS7 gateway point code and Circuit Identification Code (CIC). The SS7 DPC is associated with the Signaling Gateway that has domain over the circuit in question.
CIF	Common Intermediate Format
CIR	Committed Information Rate
CM	DOCSIS Cable Modem
CMS	Cryptographic Message Syntax
CMS	Call Management Server. Controls the audio connections. Also called a Call Agent in MGCP/SGCP terminology. This is one example of an Application Server.

CMTS	Cable Modem Termination System. The device at a cable head-end which implements the DOCSIS RFI MAC protocol and connects to CMs over an HFC network.
Codec	COder-DECoder
COPS	Common Open Policy Service Protocol. Defined in RFC 2748.
CoS	Class of Service. The type 4 tuple of a DOCSIS configuration file.
CSR	Customer Service Representative
DA	Directory Assistance
DE	Default. This is a DiffServ Per Hop Behavior.
DES	Data Encryption Standard
DHCP	Dynamic Host Configuration Protocol
DHCP-D	DHCP Default. Network Provider DHCP Server
DNS	Domain Name Service
DOCSIS®	Data Over Cable Service Interface Specifications
DPC	Destination Point Code. In ANSI SS7, a 3 octet number which uniquely identifies an SS7 Signaling Point, either an SSP, STP, or SCP.
DQoS	Dynamic Quality of Service. Assigned on the fly for each communication depending on the QoS requested.
DSCP	DiffServ Code Point. A field in every IP packet that identifies the DiffServ Per Hop Behavior. In IP version 4, the TOS byte is redefined to be the DSCP. In IP version 6, the Traffic Class octet is used as the DSCP. See [2].
DTMF	Dual-tone Multi Frequency (tones)
EF	Expedited Forwarding. A DiffServ Per Hop Behavior.
E-MTA	Embedded MTA. A single node that contains both an MTA and a cable modem.
EO	End Office
ESP	IPSec Encapsulating Security Payload. Protocol that provides both IP packet encryption and optional message integrity, not covering the IP packet header.
FEC	Forward Error Correction
ETSI	European Telecommunications Standards Institute
FEID	Financial Entity ID
FGD	Feature Group D signaling
F-link	F-Links are SS7 links that directly connect two SS7 end points, such as two SSPs. 'F' stands for "Fully Associated".
FQDN	Fully Qualified Domain Name. Refer to [9] for details.
GTT	Global Title Translation
HFC	Hybrid Fiber/Coaxial cable. An HFC system is a broadband bi-directional shared media transmission system using fiber trunks between the head-end and the fiber nodes, and coaxial distribution from the fiber nodes to the customer locations.
HMAC	Hashed Message Authentication Code. A message authentication algorithm, based on either SHA-1 or MD5 hash and defined in [10].
HTTP	Hypertext Transfer Protocol. Refer to [11] and [12].
IANA	Internet Assigned Numbered Authority. See www.ietf.org for details.
IC	Inter-exchange Carrier

ICMP	Internet Control Message Protocol. An extension to the Internet Protocol, ICMP supports packets containing error, control, and information messages.
IETF	Internet Engineering Task Force. A body responsible, among other things, for developing standards used on the Internet.
IKE	Internet Key Exchange. A key management mechanism used to negotiate and derive keys for SAs in IPSec.
IKE–	A notation defined to refer to the use of IKE with pre-shared keys for authentication.
IKE+	A notation defined to refer to the use of IKE with X509 certificates for authentication.
IP	Internet Protocol. An Internet network-layer protocol.
IPSec	Internet Protocol Security. A collection of Internet standards for protecting IP packets with encryption and authentication.
ISDN	Integrated Services Digital Network
ISP	Internet Service Provider
ISTP	Internet Signaling Transport Protocol
ISUP	ISDN User Part. A protocol within the SS7 suite of protocols that is used for call signaling within an SS7 network.
ITU	International Telecommunication Union
ITU-T	International Telecommunications Union–Telecommunications Standardization Sector
IVR	Interactive Voice Response System
KDC	Key Distribution Center
LATA	Local Access and Transport Area
LD	Long Distance
LIDB	Line Information Database. Contains information on customers required for real-time access such as calling card personal identification numbers (PINs) for real-time validation.
LLC	Logical Link Control. The Ethernet Packet header and optional 802.1P tag which may encapsulate an IP packet. A sublayer of the Data Link Layer.
LNP	Local Number Portability. Allows a customer to retain the same number when switching from one local service provider to another.
LSSGR	LATA Switching Systems Generic Requirements
MAC	Message Authentication Code. A fixed-length data item that is sent together with a message to ensure integrity, also known as a MIC.
MAC	Media Access Control. It is a sublayer of the Data Link Layer. It normally runs directly over the physical layer.
MC	Multipoint Controller
MCU	Multipoint Conferencing Unit
MD5	Message Digest 5. A one-way hash algorithm that maps variable length plaintext into fixed-length (16 byte) ciphertext.
MDCP	Media Device Control Protocol. A media gateway control specification submitted to IETF by Lucent. Now called SCTP.
MDU	Multi-Dwelling Unit. Multiple units within the same physical building. The term is usually associated with high-rise buildings
MEGACO	Media Gateway Control IETF working group. See www.ietf.org for details.

MG	Media Gateway. Provides the bearer circuit interfaces to the PSTN and transcodes the media stream.
MGC	Media Gateway Controller. The overall controller function of the PSTN gateway. Receives, controls and mediates call-signaling information between the PacketCable and PSTN.
MGCP	Media Gateway Control Protocol. Protocol follow-on to SGCP. Refer to IETF RFC [13].
MIB	Management Information Base
MIC	Message Integrity Code. A fixed-length data item that is sent together with a message to ensure integrity, also known as a Message Authentication Code (MAC).
MMC	Multi-Point Mixing Controller. A conferencing device for mixing media streams of multiple connections.
MSB	Most Significant Bit
MSO	Multi-System Operator. A cable company that operates many head-end locations in several cities.
MSU	Message Signal Unit
MTA	Multimedia Terminal Adapter. Contains the interface to a physical voice device, a network interface, Codecs, and all signaling and encapsulation functions required for VoIP transport, class features signaling, and QoS signaling.
MTP	The Message Transfer Part. A set of two protocols (MTP 2, 3) within the SS7 suite of protocols that are used to implement physical, data link and network-level transport facilities within an SS7 network.
MWD	Maximum Waiting Delay
NANP	North American Numbering Plan
NANPNAT	North American Numbering Plan Network Address Translation
NAT Network Layer	Network Address Translation. Layer 3 in the Open System Interconnection (OSI) architecture. This layer provides services to establish a path between open systems.
NCS	Network Call Signaling
NPA-NXX	Numbering Plan Area (more commonly known as area code) NXX (sometimes called exchange) represents the next three numbers of a traditional phone number. The N can be any number from 2-9 and the Xs can be any number. The combination of a phone number's NPA-NXX will usually indicate the physical location of the call device. The exceptions include toll-free numbers and ported number (see LNP)
NTP	Network Time Protocol. An internet standard used for synchronizing clocks of elements distributed on an IP network
NTSC	National Television Standards Committee. Defines the analog color television, broadcast standard used today in North America.
OID	Object Identification
OSP	Operator Service Provider
OSS	Operations Systems Support. The back-office software used for configuration, performance, fault, accounting, and security management.
OSS-D	OSS Default. Network Provider Provisioning Server
PAL	Phase Alternate Line. The European color television format that evolved from the American NTSC standard.

PCM	Pulse Code Modulation. A commonly employed algorithm to digitize an analog signal (such as a human voice) into a digital bit stream using simple analog to digital conversion techniques.
PDU	Protocol Data Unit
PHS	Payload Header Suppression. A DOCSIS technique for compressing the Ethernet, IP and UDP headers of RTP packets.
PKCROSS	Public Key Cryptography for Cross-Ream Authentication. Utilizes PKINIT for establishing the inter-realm keys and associated inter-realm policies to be applied in issuing cross-realm service tickets between realms and domains in support of Intradomain and Interdomain CMS-to-CMS signaling (CMSS).
PKCS	Public Key Cryptography Standards. Published by RSA Data Security Inc. These Standards describe how to use public key cryptography in a reliable, secure and interoperable way.
PKI	Public Key Infrastructure. A process for issuing public key certificates, which includes standards, Certification Authorities, communication between authorities and protocols for managing certification processes.
PKINIT	Public Key Cryptography for Initial Authentication. The extension to the Kerberos protocol that provides a method for using public key cryptography during initial authentication
PSC	Payload Service Class Table, a MIB table that maps RTP payload Type to a Service Class Name.
PSFR	Provisioned Service Flow Reference. An SFR that appears in the DOCSIS configuration file.
PSTN	Public Switched Telephone Network
QCIF	Quarter Common Intermediate Format
QoS	Quality of Service. Guarantees network bandwidth and availability for applications.
RADIUS	Remote Authentication Dial-In User Service. An internet protocol (IETF RFCs [14] and [15]) originally designed for allowing users dial-in access to the internet through remote servers. Its flexible design has allowed it to be extended well beyond its original intended use.
RAS	Registration, Admissions and Status. RAS Channel is an unreliable channel used to convey the RAS messages and bandwidth changes between two H.323 entities.
RC4	Rivest Cipher 4. A variable length stream cipher. Optionally used to encrypt the media traffic in PacketCable.
RFC	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.
RFI	The DOCSIS Radio Frequency Interface specification.
RJ-11	Registered Jack-11. A standard 4-pin modular connector commonly used in the United States for connecting a phone unit into a wall jack.
RKS	Record Keeping Server. The device which collects and correlates the various Event Messages.
RSA	A public-key, or asymmetric, cryptographic algorithm that is used to provide the services of authentication and encryption. RSA stands for the three inventors of the algorithm; Rivest, Shamir, Adleman.
RSA Key Pair	A public/private key pair created for use with the RSA cryptographic algorithm.
RSVP	Resource Reservation Protocol

RTCP	Real-Time Control Protocol
RTO	Retransmission Timeout
RTP	Real-time Transport Protocol. A protocol for encapsulating encoded voice and video streams. Refer to IETF RFC [16].
SA	Security Association. A one-way relationship between sender and receiver offering security services on the communication flow.
SAID	Security Association Identifier. Uniquely identifies SAs in the DOCSIS Baseline Privacy Plus Interface (BPI+) security protocol.
SCCP	Signaling Connection Control Part. A protocol within the SS7 suite of protocols that provides two functions in addition to those provided within MTP. The first is the ability to address applications within a signaling point. The second function is Global Title Translation.
SCP	Service Control Point. A Signaling Point within the SS7 network, identifiable by a Destination Point Code that provides database services to the network.
SCTP	Stream Control Transmission Protocol
SDP	Session Description Protocol
SDU	Service Data Unit. Information that is delivered as a unit between peer service access points.
SF	Service Flow. A unidirectional flow of packets on the RF interface of a DOCSIS system.
SFID	Service Flow ID. A 32-bit integer assigned by the CMTS to each DOCSIS Service Flow defined within a DOCSIS RF MAC domain. Any 32-bit SFID must not conflict with a zero-extended 14-bit SID. SFIDs are considered to be in either the upstream direction (USFID) or downstream direction (DSFID). USFIDs and DSFIDs are allocated from the same SFID number space.
SFR	Service Flow Reference. A 16-bit message element used within the DOCSIS TLV parameters of Configuration Files and Dynamic Service messages to temporarily identify a defined Service Flow. The CMTS assigns a permanent SFID to each SFR of a message.
SG	Signaling Gateway. An SG is a signaling agent that receives/sends SCN native signaling at the edge of the IP network. In particular the SS7 SG function translates variants ISUP and TCAP in an SS7-Internet Gateway to a common version of ISUP and TCAP.
SGCP	Simple Gateway Control Protocol. Earlier draft of MGCP.
SHA – 1	Secure Hash Algorithm 1. A one-way hash algorithm.
SID	Service ID. A 14-bit number assigned by a CMTS to identify an upstream virtual circuit. Each SID separately requests and is granted the right to use upstream bandwidth.
SIP	Session Initiation Protocol. An application-layer control (signaling) protocol for creating, modifying, and terminating sessions with one or more participants.
SIP+	Session Initiation Protocol Plus. An extension to SIP.
S-MTA	Standalone MTA. A single node that contains an MTA and a non-DOCSIS MAC (e.g., ethernet).
SNMP	Simple Network Management Protocol
SOHO	Small Office/Home Office

SS7	Signaling System number 7. An architecture and set of protocols for performing out-of-band call signaling with a telephone network.
SSP	Service Switching Point. SSPs are points within the SS7 network that terminate SS7 signaling links and also originate, terminate, or tandem switch calls.
STP	Signal Transfer Point. A node within an SS7 network that routes signaling messages based on their destination address. This is essentially a packet switch for SS7. It may also perform additional routing services such as Global Title Translation.
TCAP	Transaction Capabilities Application Protocol. A protocol within the SS7 stack that is used for performing remote database transactions with a Signaling Control Point.
TCP	Transmission Control Protocol
TD	Timeout for Disconnect
TFTP	Trivial File Transfer Protocol
TFTP-D	Default – Trivial File Transfer Protocol
TGS	Ticket Granting Server. A sub-system of the KDC used to grant Kerberos tickets.
TGW	Telephony Gateway
TIPHON	Telecommunications and Internet Protocol Harmonization Over Network
TLV	Type-Length-Value. A tuple within a DOCSIS configuration file.
TN	Telephone Number
ToD	Time of Day Server
TOS	Type of Service. An 8-bit field of every IP version 4 packet. In a DiffServ domain, the TOS byte is treated as the DiffServ Code Point, or DSCP.
TSG	Trunk Subgroup
UDP	User Datagram Protocol. A connectionless protocol built upon Internet Protocol (IP).
VAD	Voice Activity Detection
VBR	Variable Bit Rate
VoIP	Voice over IP

5 BACKGROUND

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

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

5.1 PacketCable Voice Communications Quality Requirements

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

5.2 Network Preparation for Codec Support

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

5.2.1 Packet Loss Control

There is a direct correlation between packet integrity and audio quality. Anecdotal codec research suggests initial 3% packet loss rate results, on average, in a reduction in Mean Opinion Score (MOS) scores of 0.5 point, on a scale of 5. Due to less-than-pristine conditions and human-detectable compromises with most codecs, the resulting audio quality for a 3% packet loss rate will be well below PSTN "toll" quality. Above 3%, codec performance falls off rapidly, and resulting voice quality is unacceptable.

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

5.2.2 Latency Control

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

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

- Audio sampling and analog-to-digital conversion
- Buffering of samples (audio framing, plus look-ahead)
- Compression processing
- Packetization of compressed data
- Local network (DOCSIS) traversal
- Routing to the backbone network
- Backbone traversal

- Far-end reception of packets and traversal of local access
- Buffering of out-of-order and delayed packets
- Decoding, decompression, and reconstruction of the audio stream

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

5.2.2.1 Latency Control: Buffering

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

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

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

5.2.2.2 Latency Control: Optimal Framing/Packetization

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

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

5.2.2.3 Latency Control: Packet Timing Optimization

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

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

Table 1. Frame Sizes of the Mandatory Codecs

Codec	Frame Size (msec)
G.711	0.125
iLBC	20
iLBC	30
BV16	5

5.2.3 Codec Transcoding Minimization

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

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

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

5.2.4 Bandwidth Minimization

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

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

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

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

6 DEVICE REQUIREMENTS FOR AUDIO CODEC SUPPORT

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

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

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

6.1 Dynamic Update Capability

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

6.2 Maximum Service Outage

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

6.3 Minimum Processing Capability

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

Table 2. MTA Processing Capability

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

6.4 Minimum Audio Codec Storage Capability

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

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

7 AUDIO CODECS SPECIFICATIONS

7.1 Feature Support

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

7.1.1 DTMF Support

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

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

The capability described above MUST be supported on all connections.

7.1.2 Fax and Modem Support

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

In order to provide customers with access for analog fax and modems, the MTA devices MUST be able to detect fax/modem signals and signal these detections using the appropriate protocol. The codec at each end is then switched to G.711 for the remainder of the session. Additionally, echo cancellation is disabled in response to a disabling signal sent by some devices (fax or modem) consisting of a 2100 Hz tone with periodic phase reversals per ITU standards G.165 [18] and G.168 [19]. After the device session has completed, echo compensation MUST be enabled.

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

MTAs and MGs MUST detect the T.30 fax preamble (V.21 flags) and CNG (calling fax tone). The detection of CNG MUST be a configurable option since it will cause calls between Super Group 3 fax machines to drop back to standard Group 3 rates (14.4kb/s max) in T.38 implementations not capable of supporting Version 3 (V.34). If CNG detection is disabled, calls between Super Group 3 fax machines will be treated as modem calls (with transmission rates of up to 33.6kb/s) as these devices do not send the T.30 fax preamble once they recognize each other through their V.8 handshaking at the start of the call. On the other hand, enabling CNG detection as a trigger to switchover to T.38 will ensure that all fax calls benefit from the use of fax relay to provide resilience from packet loss. MTAs and MGs detecting CNG MUST apply appropriate signal discrimination to minimize the chance that a voice call could inadvertently be switched to T.38 fax relay.

7.1.3 Echo Compensation Support

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

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

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

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

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

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

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

7.1.4 Asymmetrical Services Support

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

7.1.5 Hearing-impaired Services Support

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

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

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

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

7.1.6 A-law and μ-law Support

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

7.1.7 Packet Loss Concealment.

All Media Gateways and Media Terminal Adaptors MUST detect audio packet loss and implement some method to conceal losses from end-users. Specifications for low bit rate codecs (e.g. G.728, G.729, iLBC, BV16) include methods for concealment (the packet loss concealment method for iLBC, as defined and included in [34] is RECOMMENDED for iLBC and the packet loss concealment method for BV16, as defined and included in [43] is RECOMMENDED for BV16). For G.711, the method defined in ANSI T1.521-1999 is RECOMMENDED [4].

7.1.8 Fax Relay

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

The ITU-T Recommendation T.38 is a widely recognized standard for fax relay [20]. The first version for the T.38 specification is version 0 and the majority of implementations are compatible with this version, while later implementations are also required to inter-operate with version 0. MTAs and Media Gateways MUST support version 0 of the T.38 specification [20] in order to ensure interoperability with existing T.38 implementations. In addition, a MTA or a Media Gateway MAY support any other version of the T.38 specification. All client devices (MTAs and Media Gateways) MUST support the V.27ter, V.29, V.17 modem protocols for page transmission within the T.38 implementation to allow transfer rates up to 14400bps. Fax transmissions utilizing the V.34 modem protocol (super G3 fax) SHOULD be handled as described in Section 7.1.2 using the G.711 pass-through mode. However if the implementation supports T.38 Version 3 then it MAY use T.38 Version 3 mode to transport super G3 fax transmissions [20].

7.1.8.1 T.38 Over UDPTL

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

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

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

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

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

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

Table 4 shows the DQoS 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 DQoS flow characteristics described above.

7.1.8.2 T.38 Over RTP

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

7.1.9 DTMF Relay

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

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

For DTMF, PacketCable endpoints MUST support transmission and reception of RFC 2833 DTMF telephone-events 0-15 which represents the minimum level required for compliance with the RFC. PacketCable endpoints MAY support other telephone-events. If negotiated for a call, these events MUST be transferred via RFC 2833 telephony event packets regardless of the codec specified for the speech. In addition as an RTP payload type, DTMF relay MUST be secured through the PacketCable bearer encryption and authentication mechanisms defined in [38], if these are active on a call. MTAs and MGs MUST support the mandatory security options listed in [38] for DTMF relay and additionally, if the optional encryption algorithms are supported for audio codecs, then these MUST also be supported for DTMF relay.

IETF RFC 2833 [24] 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, Telcordia's LSSGR [25] 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.

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

A PacketCable endpoint MUST NOT transmit a DTMF telephone-event packet containing a duration field of value zero. A PacketCable endpoint SHOULD ignore a received DTMF telephone-event packet containing a duration field of value zero.

The repetition rate of RFC2833 telephony event packets in the transmit direction MUST be equal to the same packetization time as the selected audio codec. Therefore the repetition rate of RFC2833 packets has the same range as packetization intervals, i.e. 10, 20, and 30ms.

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

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

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

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

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

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

7.2 Mandatory Codecs

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

7.2.1 G.711

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

7.2.2 iLBC

iLBC was selected by CableLabs to be one of the PacketCable 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 work [34], [35]. Experimental track IETF RFC "internet Low Bit Rate Codec (iLBC)", [35] contains the iLBC source code in floating point C. It was created to provide a codec suitable for IP communication networks and is available on a Royalty Free basis.

iLBC MUST be supported in all MTAs and MGs. PacketCable has as a mandate to provide toll or superior voice quality. iLBC is a mid-bitrate (13.3 kb/s and 15.2 kb/s), high-quality solution. When iLBC is supported, both the 20ms and 30ms frame size modes MUST be supported. Mandating a codec in this range provides high quality, low-bandwidth performance and high packet loss robustness for on-net calls and ensures high performance for applications such as IVR systems. In addition, it provides DTMF pass through.

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

7.2.3 BV16

BroadVoice16 [37] [43] MUST be supported in all MTAs and MGs. PacketCable has as a mandate to provide toll or superior voice quality. BroadVoice16 (BV16) is a mid-bitrate, high-quality solution. Mandating a codec in this range provides high quality, low-bandwidth performance for on-net calls and ensures high performance for applications such as IVR systems. In addition, it provides DTMF pass through. It was created to provide a codec suitable for IP communication networks and is available on a Royalty Free basis for PacketCable.

7.3 Recommended Codecs

In addition to G.711, iLBC and BV16, it is RECOMMENDED that MTAs and MGs also support at least one of the following codecs.

7.3.1 G.728

G.728 [22] SHOULD be supported in all MTAs. PacketCable has as a mandate to provide toll or superior voice quality. G.728 is a mid-bitrate (16 kb/s), high-quality solution. Mandating a codec in this range provides high quality, low-bandwidth performance for on-net calls and ensures the highest possible performance for applications such as IVR systems. In addition, it provides superior background noise handling, as well as medium quality music carriage.

7.3.2 G.729 Annex E

G.729 Annex E [23] SHOULD be supported in all MTAs. PacketCable has as a mandate to provide toll or superior voice quality. G.729E is a mid-bitrate (11.8 kb/s), high-quality solution. Mandating a codec in this range provides high quality, low-bandwidth performance for on-net calls and ensures the highest possible performance for applications such as IVR systems. In addition, it provides superior background noise handling, as well as medium quality music carriage.

7.4 Optional Features

7.4.1 Wideband Codecs

Given that the majority of early customers will be "black phone" users, support for wideband (i.e., greater than circuit voice bandwidth) codecs is not being mandated. However, some vendors optionally MAY choose to differentiate their product by selecting components that will support higher fidelity in the event a wideband codec is provisioned through methods specified in Section 6.1.

7.4.2 Optional Codecs

A vendor MAY supply any codecs not described herein.

7.4.3 Voice Activity Detection (VAD)

A vendor MAY employ VAD to reduce bandwidth consumption. If employed, this capability MUST be optional, allowing disabling. Some codecs have associated VAD implementations (e.g. G.729B), while many others do not (e.g. G.711 and G.728). In the latter cases, the VAD implementation MUST adhere to the IMTC Voice-Over-IP Forum Service Interoperability Implementation Agreement 1.0 [5].

7.5 Session Description of Codecs

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

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

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

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

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

a=<token>:<value>

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

m=audio 3456 RTP/AVP 0

a=ptime:10

On the transport address line (m), the first term defines the media type, which in the case of an IP voice communications session is audio. The second term defines the UDP port to which the media is sent (port 3456). The third term indicates that this stream is an RTP Audio/Video profile. Finally, the last term is the media payload type as defined in the RTP Audio/Video Profile, IETF RFC 3551, [7]. 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 [7] are dynamically bound by using a dynamic payload type from the range 96-127, as defined in [8], and a media attribute line. For example, a typical SDP message for G.726 would be composed as follows:

m=audio 3456 RTP/AVP 96

a= rtpmap:96 G726-32/8000

The payload type 96 indicates that the payload type is locally defined for the duration of this session, and the following line indicates that payload type 96 is bound to the encoding "G726-32" with a clock rate of 8000 samples/sec.

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

Table 3. Codec RTP Map Parameters

Codec	Literal Codec Name	RTP Map Parameter
G.711 μ-law	PCMU	PCMU/8000
G.711 A-law	PCMA	PCMA/8000
iLBC	iLBC	iLBC/8000
BroadVoice16	BV16	BV16/8000

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

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

For every defined Codec (whether it is represented in SDP as a static or dynamic payload type), the following table 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 4 do not include any bandwidth that may be required for media security (authentication, 2 or 4 byte value as outlined in the security specification), and the actual values used in resource allocation may need to be adjusted to accommodate 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 in the SDP, the bandwidth parameter should contain the least-upper-bound (LUB) of the desired codec bandwidths.

The mapping of RTP/AVP code to RSVP Flowspec (as used by Dynamic Quality of Service [2]) MUST be according to the following table:

Table 4. Mapping of Session Description Parameters to RSVP Flowspec

Parameters from Session Description			Flowspec parameters		Comments
RTP/AVP code	Rtpmap	Ptime (msec)	Values b,m,M[1]	Values r,p[2]	
0	<none>	10	120 bytes	12,000 bytes/sec	G.711 µ-law using the Payload Type defined by IETF
0	<none>	20	200 bytes	10,000 bytes/sec	
0	<none>	30	280 bytes	9,334 bytes/sec	
96-127	PCMU/8000	10	120 bytes	12,000 bytes/sec	G.711 µ-law PCM, 64 kb/sec, default Codec
96-127	PCMU/8000	20	200 bytes	10,000 bytes/sec	
96-127	PCMU/8000	30	280 bytes	9,334 bytes/sec	
8	<none>	10	120 bytes	12,000 bytes/sec	G.711 A-law using the Payload Type defined by IETF
8	<none>	20	200 bytes	10,000 bytes/sec	
8	<none>	30	280 bytes	9,334 bytes/sec	
96-127	PCMA/8000	10	120 bytes	12,000 bytes/sec	G.711 A-law PCM, 64 kb/sec, default Codec
96-127	PCMA/8000	20	200 bytes	10,000 bytes/sec	
96-127	PCMA/8000	30	280 bytes	9,334 bytes/sec	
96-127	iLBC/8000	20	78 bytes	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	G726-16/8000	10	60 bytes	6,000 bytes/sec	
96-127	G726-16/8000	20	80 bytes	4,000 bytes/sec	
96-127	G726-16/8000	30	100 bytes	3,334 bytes/sec	
96-127	G726-24/8000	10	70 bytes	7,000 bytes/sec	
96-127	G726-24/8000	20	100 bytes	5,000 bytes/sec	
96-127	G726-24/8000	30	130 bytes	4,334 bytes/sec	
2	<none>	10	80 bytes	8,000 bytes/sec	G.726-32, identical to G.721, which is assigned Payload Type 2 by IETF
2	<none>	20	120 bytes	6,000 bytes/sec	
2	<none>	30	160 bytes	5,334 bytes/sec	
96-127	G726-32/8000	10	80 bytes	8,000 bytes/sec	
96-127	G726-32/8000	20	120 bytes	6,000 bytes/sec	
96-127	G726-32/8000	30	160 bytes	5,334 bytes/sec	

[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).

Parameters from Session Description			Flowspec parameters		Comments
RTP/AVP code	**Rtpmap**	**Ptime (msec)**	**Values b,m,M[1]**	**Values r,p[2]**	
96-127	G726-40/8000	10	90 bytes	9,000 bytes/sec	
96-127	G726-40/8000	20	140 bytes	7,000 bytes/sec	
96-127	G726-40/8000	30	190 bytes	6,334 bytes/sec	
15	<none>	10	60 bytes	6,000 bytes/sec	G.728, assigned Payload Type 15 by IETF
15	<none>	20	80 bytes	4,000 bytes/sec	
15	<none>	30	100 bytes	3,334 bytes/sec	
96-127	G728/8000	10	60 bytes	6,000 bytes/sec	G.728, LD-CELP, 16kb/s
96-127	G728/8000	20	80 bytes	4,000 bytes/sec	
96-127	G728/8000	30	100 bytes	3,334 bytes/sec	
18	<none>	10	50 bytes	5,000 bytes/sec	G.729A, identical to G.729, assigned Payload Type 18 by IETF
18	<none>	20	60 bytes	3,000 bytes/sec	
18	<none>	30	70 bytes	2,334 bytes/sec	
96-127	G729/8000	10	50 bytes	5,000 bytes/sec	G.729A, CS-ACELP, 8kb/s, 10ms frame size with 5ms lookahead
96-127	G729/8000	20	60 bytes	3,000 bytes/sec	
96-127	G729/8000	30	70 bytes	2,334 bytes/sec	
96-127	G729E/8000	10	55 bytes	5,500 bytes/sec	G.729E, CS-ACELP, 11.8kb/s, 10ms frame size with 5ms lookahead
96-127	G729E/8000	20	70 bytes	3,500 bytes/sec	
96-127	G729E/8000	30	85 bytes	2,834 bytes/sec	
N/A	N/A	10	80 bytes	8000 bytes/sec	T.38 fax relay packets (with T.4 redundancy level 1, T30 redundancy level 4)
N/A	N/A	20	116 bytes	5800 bytes/sec	
N/A	N/A	30	152 bytes	5067 bytes/sec	
Table Note: Mandatory codecs – G.711 (μ-law and A-law), iLBC, BV16 Recommended codecs – G.728 and G.729 Annex E Optional codecs (for informational purposes only) – G.726 and G.729A					

7.5.1 iLBC Session Description

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

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

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

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

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

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

7.5.2 BV16 Session Description

Parameters are mapped to SDP in a standard way. When conveying information by SDP, the encoding name MUST be "BV16" (the same as the MIME subtype [43]).

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

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

8 VIDEO REQUIREMENTS

8.1 Overview

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

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

This section addresses details of video communication over the PacketCable network—in particular, the video codec requirements. The H.261 [47] and H.263 [48] Recommendations (as well as H.245 [46], or a functionally equivalent specification) are the basis and reference for this specification; highlights of these recommendations important to PacketCable are illustrated here. Additionally, issues that have dependencies upon other PacketCable resources, such as signaling and quality-of-service (QoS), are outlined.

8.2 PacketCable Video Devices

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

8.3 Video Encoder Requirements

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

For reasons of interoperability, all PacketCable MTA-2 terminals providing video communications MUST be capable of encoding and decoding video according to H.261. This will permit video communication without the transcoding of video with terminals across the other networks, such as H.320 [49] terminals across an ISDN network or an H.324 [50] terminal across a PSTN network. The use of H.261 establishes a common denominator across all communication networks and retains backward compatibility with existing systems.

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

8.4 Video Format Requirements

As stated in the H.263 recommendation:

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

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

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

Picture Format	Number of pixels for luminance (dx)	Number of lines for luminance (dy)	Number of pixels for chrominance (dx/2)	Number of lines for chrominance (dy/2)
SQCIF	128	96	64	48
QCIF	176	144	88	72
CIF	352	288	176	144
4CIF	704	576	352	288
16CIF	1408	1152	704	576

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

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

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

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

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

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

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

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

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

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

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

8.5 H.263 Annexes

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

Annex D. Unrestricted Motion Vector Mode

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

Recommendation: MTA-2s SHOULD employ this mode.

Annex E. Syntax-based Arithmetic Coding

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

Recommendation: MTA-2s SHOULD NOT employ this mode.

Annex F. Advanced Prediction Mode

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

Recommendation: MTA-2s SHOULD employ this mode.

Annex G. PB-Frames Mode

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

Recommendation: MTA-2s SHOULD NOT employ this mode.

Annex H. Forward Error Correction for Coded Video Signal

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

Recommendation: MTA-2s SHOULD NOT employ this mode.

Annex I. Advanced INTRA Coding Mode

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

Recommendation: MTA-2s SHOULD employ this mode.

<u>*Annex J. Deblocking Filter Mode*</u>

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

Recommendation: MTA-2s SHOULD employ this mode.

<u>*Annex K. Slice Structured Mode*</u>

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

Recommendation: MTA-2s SHOULD employ this mode.

<u>*Annex L. Supplemental Enhancement Information*</u>

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

Recommendation: MTA-2s SHOULD employ this mode.

<u>*Annex M: Improved PB-Frames Mode*</u>

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

Recommendation: MTA-2s SHOULD NOT employ this mode.

<u>*Annex N: Reference Picture Selection Mode*</u>

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

Recommendation: MTA-2s MAY employ this mode.

<u>*Annex O. Temporal/SNR/Spatial Scalability*</u>

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

Recommendation: MTA-2s SHOULD NOT employ this mode.

<u>*Annex P. Reference Picture Resampling*</u>

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

Recommendation: MTA-2s SHOULD NOT employ this mode.

<u>*Annex Q. Reduced Resolution Update Mode*</u>

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

Recommendation: MTA-2s SHOULD NOT employ this mode.

<u>*Annex R. Independently Segmented Decoding Mode*</u>

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

Recommendation: MTA-2s SHOULD employ this mode.

Annex S. Alternative INTER VLC Mode

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

Recommendation: MTA-2s SHOULD NOT employ this mode.

Annex T. Modified Quantization Mode

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

Recommendation: MTA-2s SHOULD employ this mode.

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

Table 6. H.263 Annexes and their Applicability to PacketCable

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

8.6 Multipoint Conferencing Support

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

8.6.1 Freeze Picture Request

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

8.6.2 Fast Update Request

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

8.6.3 Freeze Picture Release

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

8.6.4 Continuous Presence Multipoint (CPM)

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

8.7 Signaling Messages

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

H.245 [46] provides an example of essential signaling components vital to an MTA-2 video call. Not only can H.245 be used for the exchange of capabilities at the initialization of a call, it may also be used during a call for several video and conference-centric commands. A list of mandatory (M) and optional (O) signals from the H.245 command set is shown below for receiving and transmitting MTA-2s. The mandatory commands (or their functional equivalents) MUST be implemented in the PacketCable signaling system.

Table 7. H.245 Commands that are Applicable to PacketCable

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

Table Note:
M = mandatory
O = optional

9 RTP AND RTCP USAGE

9.1 RTP Requirements

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

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

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

9.2 RTCP Requirements

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

9.2.1 General Requirements of the PacketCable RTCP Profile

PacketCable endpoints MUST send RTCP messages, as described in RFCs [16] and [7] and profiled below.

Endpoints MAY start transmitting RTCP messages as soon as the RTP session has been established, even if RTP packets are not being sent or received. An RTP session is considered established once each endpoint has received a remote connection descriptor. Furthermore, 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 unnecessary network traffic, endpoints MAY stop sending RTCP packets to a remote endpoint if an ICMP port unreachable or another ICMP destination unreachable error (i.e., ICMP error type 3) is returned from the network for that RTCP destination.

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

An RTCP transmission interval calculation procedure is outlined in Section 9.2 of this document.

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

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

SSRC (Synchronization Source) collision detection and resolution is OPTIONAL for PacketCable endpoints that are capable of unambiguously distinguishing between media packets and reports that they send and those that it receives. If an endpoint can handle SSRC collisions without affecting the integrity of the session, the endpoint MAY ignore SSRC collisions. In particular, SSRC collision detection and resolution is OPTIONAL for endpoints that are establishing unicast, point-to-point connections carrying one RTP stream, as is the case in current PacketCable connections. If SSRC collision detection and resolution is supported, one or both of the endpoints MUST resolve SSRC collisions as follows: (1) send BYE, (2) select new SSRC, (3) send Sender Description with new SSRC. SSRC collision detection and resolution is OPTIONAL for PacketCable endpoints that perform mixing for multiple remote endpoints when CSRC lists are not transmitted in the mixed packets. When CSRC lists are transmitted, the mixing endpoint MUST detect and resolve SSRC collisions.

Future PacketCable connections may involve multiple, simultaneous RTP streams, and require resolution of SSRC collisions. In this case responsibility for this resolution falls to the two colliding senders. One or both of these parties MUST resolve SSRC collisions as follows: (1) send BYE, (2) select new SSRC, (3) send Sender Description with new SSRC.

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

SDES (Source Description): CNAME objects MUST NOT contain identity information (see definition below); CNAME field MUST be a cryptographically-random value generated by the endpoint in such a manner that endpoint identity is not compromised and MUST change on a per-session basis; NAME, EMAIL, PHONE, LOC objects SHOULD NOT be sent and, if sent, MUST NOT contain identity information. This requirement is intended to satisfy the requirements of [16] with respect to the CNAME field, and at the same time satisfy legal and regulatory requirements for maintaining subscriber privacy, for example, when caller id blocking must be performed. This requirement is imposed because not all RTCP messages may be encrypted, as described in the PacketCable Security Specification [38].

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

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

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

BYE (Goodbye): MUST be sent upon RTP connection deletion or when renegotiating SSRC upon collision detection and resolution (see below). Endpoints MUST send BYE commands when the application needs to discontinue use of an SSRC and start a new SSRC, for example, on codec change. (Note: codec change is an example only, since in some implementations, the endpoint may not need to change SSRC when changing codec.) Endpoints MUST NOT use BYE messages to indicate or detect any call progress condition. For example, endpoints MUST NOT tear down RTP flows based on BYE, but MUST update RTCP/RTP state as per RFC [16]. This requirement is intended to ensure that all call progress conditions, such as on-hook notifications, are signaled using the higher-level PacketCable signaling protocol, such as Network-based Call Signaling (NCS).

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.

9.2.2 Security Requirements for RTP and RTCP in PacketCable

PacketCable endpoints MUST NOT conform to the security requirements described in the RTP/RTCP RFC and drafts. Instead, PacketCable endpoints MUST implement RTP and RTCP security as specified in the PacketCable Security Specification [38].

9.2.3 Extended RTCP Reports

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

9.2.3.1 Reporting Call Quality Metrics using RTCP XR

9.2.3.1.1 RTCP XR VoIP Metrics Requirements

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

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

PacketCable endpoints that support the RTCP XR VoIP Metrics payload MUST measure or compute the reported values of the metrics as defined in Section 9.2.3.1.2 to 9.2.3.1.6 of this Specification.

9.2.3.1.2 Definition of Metrics related to Packet Loss and Discard

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

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

Table 8. Metrics Related to Packet Loss and Discard

METRIC	Description	Range
Loss Rate	Proportion of packets lost within the network	0 to 0.996
Discard Rate	Proportion of packets discarded due to late arrival	0 to 0.996
Burst Loss Density	Proportion of packets lost and discarded during burst periods	0 to 0.996
Gap Loss Density	Proportion of packets lost and discarded during gap periods	0 to 0.996
Burst Duration	Average length of burst periods (ms)	0 to 65,535
Gap Duration	Average length of gap periods (ms)	0 to 65,535
Gmin	Parameter used to define burst periods	16

A PacketCable endpoint when using RTCP XR MUST provide these parameters as defined in Table 8.

9.2.3.1.3 Definition of Metrics Related to Delay

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

Table 9. Metrics related to Delay

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

A PacketCable endpoint when using RTCP XR MUST provide the parameters as defined in Table 9. [Note that this requires an SR or RR exchange prior to the inclusion of an XR payload into an RTCP message].

9.2.3.1.4 Definition of Metrics related to Signal

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

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

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

Table 10. Metrics due to Signal

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

A PacketCable endpoint when using RTCP XR MUST provide Signal Level and Noise as defined in Table 10.

A PacketCable endpoint equipped with an echo canceller and when using RTCP XR MUST provide the Residual Echo Return Loss metric as defined in Table 10.

9.2.3.1.5 Definition of Metrics related to Call Quality

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

The maximum range of an R factor is 0-100 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:

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

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

Strictly, a MOS can only be obtained from subjective testing, however the MOS scale represents a convenient and well understood scale, and hence is often used. ITU G.107 [32] 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 11. Metrics related to Call Quality

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

A PacketCable endpoint when using RTCP-XR MUST provide the R Factor, MOS-LQ and MOS-CQ metrics and MAY provide an External R Factor.

A PacketCable endpoint when using RTCP XR MUST calculate R Factors using G.107 at a minimum [32].

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

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

E Model No parameter = Noise Level

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

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

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

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

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

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

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

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

E Model Ta = T = Total Delay / 2

E Model Tr = Total Delay

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

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

A PacketCable endpoint when using RTCP XR MUST calculate the Ie,eff parameter using the function defined in G.107 [32]. However, the PacketCable endpoint MUST use the Ie and Bpl parameters defined in Table 12 for the Vocoder and PLC combinations listed.

Table 12. Ie and Bpl parameters for PacketCable Vocoders

Vocoder	Bit rate	PLC	Ideal R	Ideal MOS	Ie	Bpl
G.711 A/U	64k	Appendix 1 [4]	93	4.4	0	34
G.728 10ms	16k	Per G.728 Annex I [22]	89	4.1	7	17
G.728 20ms	16K	Per G.728 Annex I [22]	89	4.1	7	15
G.729 Annex E 10ms	11.8k	Per G.729 [23]	88	4.1	4	20
G.729 Annex E 20ms	11.8K	Per G.729 [23]	88	4.1	4	19
ILBC 20mS	15.2k	Per [36]	80	3.9	10	34
ILBC 30mS	13.3k	Per [36]	78	3.8	12	27
BV16 10ms	16k	Per [43]	88	4.2	5	25
BV16 20ms	16K	Per [43]	88	4.2	5	23

A PacketCable endpoint when using RTCP XR MUST calculate MOS-LQ using the R to MOS mapping function defined in G.107 [32] applied to the value (R – Id).

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

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

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

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

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

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

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 [33] score, gives the closest match to the measured P.862 [33] score. Alternatively, the Bpl value may be obtained by comparing the P.862 [33] score curve with other codecs with known Bpl factor.

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

9.2.3.1.6 Definition of Parameters related to endpoint configuration

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

Table 13. Parameters related to endpoint configuration

METRIC	Description	Range
PLC Type	Type of packet loss concealment algorithm:	UnspecifiedDisabled EnhancedStandard
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

A PacketCable endpoint when using RTCP XR MUST provide values to all Parameters as defined in Table 13.

Appendix A. Codec Comparison Tables

The following three tables summarize standard speech coder characteristics.

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

Table 14. ITU IETF and CableLabs Speech Coders

Standards Body	ITU	ITU	ITU	ITU	ITU	ITU	ITU	ITU	IETF[1]/ Cable Labs	Cable Labs
Recom-mendation	G.711	G.726	G.728	G.729	G.729A	G.729D	G.729E	G.723.1	iLBC	BV16
Coder Type	Companded PCM	ADPCM	LD-CELP	CS-ACELP	CS-ACELP	CS-ACELP	CS-ACELP	MPC-MLQ & ACELP	FB-LPC	TSNFC
Dates	1972	1990	1992/4	1995	1996	1998	1998	1995	2002	2003
Bitrate	64 kb/s	16-40 kb/s	16 kb/s	8 kb/s	8 kb/s	6.4 kb/s	11.8 kb/s	6.3 kb/s & 5.3 kb/s	15.2kb/s & 13.3 kb/s	16 kb/s
Peak Quality[2]	Toll	≤Toll	Toll	Toll	Toll	< Toll	Toll	≤Toll	Toll	Toll
Background Noise[3]	Toll	≤ Toll	Toll	≤ Toll	≤ Toll	< Toll	Toll	≤ Toll	Toll	Toll
Tandem[4]	Toll	Toll	Toll	< Toll	< Toll	< Toll	Toll	< Toll	< Toll	Toll
Frame Erasure[5]	No mechanism	No mechanism	3%	3%	3%	3%	3%	3%	7% & 5%	5%
Complexity (MIPS)[6]	~0.35	~12	~36	~22	~13	~20	~27	~19	~15 &~18	~12
RAM (kword)[7]	~0.01	~0.15	~2.20	~2.6	~2.6	~2.6	~2.6	~2.1	~4	~2
Frame Size	0.125 ms	0.125 ms	0.625 ms	10 ms	10 ms	10 ms	10 ms	30 ms	20ms & 30 ms	5 ms
Look Ahead	0	0	0	5 ms	5 ms	5 ms	5 ms	7.5 ms	5ms & 10 ms	0
Codec Delay[8]	0.25 ms	0.25 ms	1.25 ms	25 ms	25 ms	25 ms	25 ms	67.5 ms	45ms & 70 ms	10 ms

Table Notes:

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

Table 15. North American Wireless Speech Coders

Standards Body	TIA	TIA	TIA	TIA	TIA	ETSI	ETSI	ETSI
Recom-mendation	IS-54	IS-641	IS-96	IS-127	IS-733	GSM-(FR)	GSM-(HR)	GSM-(EFR)
System	TDMA	TDMA	CDMA	CDMA	CDMA	GSM	GSM	GSM
Coder Type	VSELP	ACELP	QCELP	ACELP	CELP	RPE-LTP	VSELP	ACELP
Dates	1990	1995	1993	1997	1997	1987	1994	1995
Bitrate	7.95 kb/s	7.4 kb/s	0.8-8.0 kb/s	0.8-8.55 kb/s	0.8-13.2 kb/s	13 kb/s	5.6 kb/s	12.2 kb/s
Peak Quality[1]	= GSM-(FR)	Toll	= GSM-(FR)	Toll	Toll	<Toll	=GSM-(FR)	Toll
Background Noise[2]	<< Toll	< Toll	<< Toll	< Toll	Toll	<Toll	< GSM-(FR)	Toll
Tandem[3]	<< Toll	< Toll	<< Toll	< Toll	Toll	<< Toll	< GSM-(FR)	Toll
Frame Erasures[4]	3%	3%	3%	3%	3%	3%	3%	3%
Complexity (MIPS)[5]	~12	~15	~18	~25	~22	~5	~24	~18
RAM (kword)[6]	~1.5	~2.5	~2	~2.5	~2.5	~1	~4	~4.6
Frame Size	20 ms	20 ms	20 ms	20 ms	20 ms	20 ms	20 ms	20 ms
Look Ahead	5 ms	5 ms	5 ms	5 ms	5 ms	0	4.4 ms	0
Codec Delay[7]	45 ms	45 ms	45 ms	45 ms	45 ms	40 ms	44.4 ms	40 ms

Table Notes:

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

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

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

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

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

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

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

G.729 was finalized in 1995 originally by the ITU to be a toll quality 8 kb/s standard. In that year, the ITU was requested to create a low-complexity coder for simultaneous voice and data. G.729A was created as a low-complexity version that is fully interoperable with G.729. G.729B is a speech/silence detector and comfort noise generator. It can be used with either G.729 or G.729A to provide an option for variable rate usage, also known as discontinuous transmission. G.729C contains the floating-point versions of G.729 and G.729A. G.729D is a 6.4 kb/s version of G.729. It was created to provide an optional lower rate that can be used briefly for periods of network congestion, or when more bits are needed for channel error protection. Its quality is less than that of G.729 or G.729A. G.729E is a higher rate version of G.729 designed to provide higher quality for background noise conditions, music, and tandems. It is a hybrid coder. It codes each frame two different ways and selects the method that appears to give the greater fidelity. Its forward-adaptive mode uses CS-ACELP. Its backward-adaptive mode features a 30th-order backward-adaptive LPC synthesis filter and no pitch predictor. This mode is better for music, and it has greater complexity than the original G.729 coders.

Table 16 is intended to provide essential access network bandwidth-related information for each codec listed. Although some of the listed codecs (e.g., G.711, G.726) are sample-based rather than frame-based, for anticipated purposes of flow management, frame-oriented packet sizes are listed. The three most important packet sizes are shown, corresponding to low latency (10, 20, and 30 ms) samples. Packet header overhead is calculated at 40 bytes, with 12 bytes RTP, 8 bytes UDP, and 20 bytes IP contributions. Note that G.729E is shown at a byte-boundary 12 kb/s, which includes the 2 bits/frame not currently defined. Variable bit rate VAD implementations for each codec are not listed.

Table 16. Bandwidth Attributes of Codecs

Codec	Bitrate (kb/s)	Byte/10 ms	Frm/Pkt	Byte/Pkt	Pkt/s	Byte/s	kb/s
G711-10ms	64	80	1	120	100	12000	96
G711-20ms	64	80	2	200	50	10000	80
G711-30ms	64	80	3	280	33.3	9333	75
G.726.16-10ms	16	20	1	60	100	6000	48
G.726.16-20ms	16	20	2	80	50	4000	32
G.726.16-30ms	16	20	3	100	33.3	3333	27
G.726.24-10ms	24	30	1	70	100	7000	56
G.726.24-20ms	24	30	2	100	50	5000	40
G.726.24-30ms	24	30	3	130	33.3	4333	35
G.726.32-10ms	32	40	1	80	100	8000	64
G.726.32-20ms	32	40	2	120	50	6000	48
G.726.32-30ms	32	40	3	160	33.3	5333	43
G.726.40-10ms	40	50	1	90	100	9000	72
G.726.40-20ms	40	50	2	140	50	7000	56
G.726.40-30ms	40	50	3	190	33.3	6333	51
G.728-10ms	16	20	1	60	100	6000	48
G.728-20ms	16	20	2	80	50	4000	32
G.728-30ms	16	20	3	100	33.3	3333	27
G.729A-10ms	8	10	1	50	100	5000	40
G.729A-20ms	8	10	2	60	50	3000	24
G.729A-30ms	8	10	3	70	33.3	2333	19
G.729E-10ms	12	15	1	55	100	5500	44
G.729E-20ms	12	15	2	70	50	3500	28
G.729E-30ms	12	15	3	85	33.3	2833	23
iLBC-20ms	15.2	19	1	78	50	3900	31
iLBC-30ms	13.3	16.67	1	90	33.3	3000	24
BV16-10ms	16	20	2	60	100	6000	48
BV16-20ms	16	20	4	80	50	4000	32
BV16-30ms	16	20	6	100	33.33	3333	26.7

Appendix B. Acknowledgements

This document was developed and influenced by numerous individuals representing many different organizations. PacketCable wishes to thank everybody who participated directly or indirectly in this effort. In particular, PacketCable wants to recognize the following individuals for their significant involvement and contributions to this document.

Mike Noonen (8x8)

Richard Cox (AT&T)

Donnie Tew (Arris)

Chad Griffiths, Raymond Chen (Broadcom)

Ram Jagadeesan, Mike Knappe, Bill Foster (Cisco)

Roar Hagen (Global IP Sound AB)

Dennis Ng, Vince Rhee, Phillip Freyman (Motorola)

Scott McNeill, Harpit Chhatwal (Nuera)

Alan Clark (Telchemy)

David Lide, Dan Thomas (Telogy)

Satish Kumar (Texas Instruments)

Venkatesh Sunkad (CableLabs)