MPLS Forum

Voice over MPLS – Bearer Transport

Implementation Agreement

MPLS Forum 1.0

MPLS Forum Technical Committee
July 27, 2001

Note: The user's attention is called to the possibility that implementation of the MPLS implementation agreement contained herein may require the use of inventions covered by patent rights held by third parties. By publication of this MPLS implementation agreement the MPLS Forum makes no representation that the implementation of the specification will not infringe on any third party rights. The MPLS Forum take no position with respect to any claim that has been or may be asserted by any third party, the validity of any patent rights related to any such claims, or the extent to which a license to use any such rights may not be available.

Editors:

Rao Cherukuri
Integral Access

Tom Walsh
Lucent Technologies

For more information contact:

The MPLS Forum
Suite 307
39355 California Street
Fremont, CA 94538 USA

Phone: +1 (510) 608-3997
FAX: +1 (510) 608-5917
E-Mail: info@mplsforum.org
WWW: http://www.mplsforum.org/

Full Notice

Table of Contents

1 Introduction

1.1 Purpose

The purpose of this Implementation Agreement (IA) is to define a method for conveying voice directly over MPLS without first encapsulating the voice sample in IP. There are many possible arrangements in which voice may be carried in an MPLS environment. Two of the most commonly discussed arrangements are:

- Voice over IP (VoIP) over MPLS. In this case, the typical protocol stack contains voice samples encapsulated in IP layer protocols (e.g., RTP/UDP/IP) followed by encapsulation in the MPLS protocol. Compressed headers may be utilized in some implementations. The result is then conveyed by an MPLS transport arrangement such as Frame Relay, ATM, PPP, or Ethernet.

- Voice directly over MPLS (without the IP encapsulation of the voice packet). In this case, the typical protocol stack would consist of voice samples encapsulated in the MPLS protocol on top of an MPLS transport arrangement such as Frame Relay, ATM, PPP, or Ethernet.

The first arrangement, Voice over IP (VoIP) over MPLS, is essentially a method of implementing VoIP and is largely supported by existing IETF standards. VoIP over MPLS (VoIP/MPLS) is not the subject or purpose of this implementation agreement.

The second arrangement, Voice directly over MPLS (VoMPLS), provides a very efficient transport mechanism for voice in the MPLS environment and is the subject and purpose of this implementation agreement. There are many similarities to this arrangement and other architectures in use today for Voice over ATM (VoATM) and Voice over Frame Relay (VoFR).

The purpose of this VoMPLS – Bearer Transport Implementation Agreement is to define how a voice payload is encapsulated directly in the MPLS frame. It includes the definition of a VoMPLS header format supporting various payload types including Audio, Dialed digits (DTMF), Channel Associated Signaling and a Silence insertion descriptor. The defined VoMPLS – Bearer Transport header formats are different from RTP formats that are used in Voice over IP.

1.2 Scope and Overview

This specification defines MPLS support for the transport of digital voice payloads. Frame formats and procedures required for voice transport are described in this Implementation Agreement. This specification addresses the following functions:

- Transport of uncompressed (i.e., G.711 64 kbps) and compressed voice within the payload of an MPLS frame. Support for a diverse set of voice compression algorithms;
- Silence removal and Silence insertion descriptors;
- Dialed digits (DTMF information); and
- Channel associated signaling bits.

This Implementation Agreement does not define algorithms for encoding audio streams. It references existing algorithms and specifies how the bits that they output are conveyed within an MPLS packet structure. Support for the unique needs of the different voice compression algorithms is accommodated with algorithm-specific "transfer syntax" definitions. These definitions establish algorithm specific frame formats and procedures.

Transport of supporting information for voice communication, such as signaling indications (e.g., ABCD bits), and dialed digits, is also provided through the use of transfer syntax definitions specific to the information being sent.

This Implementation Agreement does not specify signaling protocols, call routing, equipment aspects, performance guidelines, or implementation techniques. In this document, VoMPLS shall refer only to the arrangement of Voice (without IP encapsulation) over MPLS.

1.3 Definitions

Must, Shall or Mandatory — the item is an absolute requirement of this implementation agreement.

Should — the item is desirable.

May or Optional — the item is not compulsory, and may be followed or ignored according to the needs of the implementer.

Notes — outside of Tables and Figures are informative.

1.4 Acronyms

AAL	ATM Adaptation Layer
ADPCM	Adaptive Differential Pulse Code Modulation
AMR	Adaptive Multi-Rate
ATM	Asynchronous Transfer Mode
CAS	Channel Associated Signaling
CID	Channel Identification
CR-LDP	Constraint-based Routing Label Distribution Protocol
CS-ACELP	Conjugate Structure – Algebraic Code Excited Linear Predictive
DTMF	Dual Tone Multi-Frequency
E-ADPCM	Embedded Adaptive Differential Pulse Code Modulation
FQC	Frame Quality Classification
GW	Gateway
IA	Implementation Agreement
IP	Internet Protocol
ISDN	Integrated Services Digital Network
LDP	Label Distribution Protocol
LER	Label Edge Router
LPC	Linear Predictive Coding
LSB	Least Significant Bit
LSP	Label Switched Path
LSR	Label Switching Router
MPLS	Multi Protocol Label Switching

MP-MLQ	Multi Pulse Maximum Likelihood Quantizer
MSB	Most Significant Bit
PCM	Pulse Code Modulation
PDL	PAD Length
PPP	Point to Point Protocol
PSTN	Public Switched Telephone Network
RSVP-TE	Resource Reservation Protocol – Traffic Engineering
RTP	Real-time Transport Protocol
SID	Silence Insertion Descriptor
UDP	User Datagram Protocol
VAD	Voice Activity Detection
VoATM	Voice over ATM
VoFR	Voice over Frame Relay
VoIP	Voice over IP
VoMPLS	Voice over MPLS

1.5 Normative References

[1] ITU-T Recommendation G.711 — Pulse Code Modulation of Voice Frequencies, 1988

[2] ITU-T Recommendation G.723.1 — Dual Rate Speech Coder for Multimedia Communications Transmitting at 5.3 & 6.3 kbit/s, March 1996

[3] ITU-T Recommendation G.726 — 40, 32, 24, 16 kbit/s Adaptive Differential Pulse Code Modulation (ADPCM), March 1991

[4] ITU-T Recommendation I.366.2 — AAL Type 2 Service Specific Convergence Sublayer for Narrow-band Services

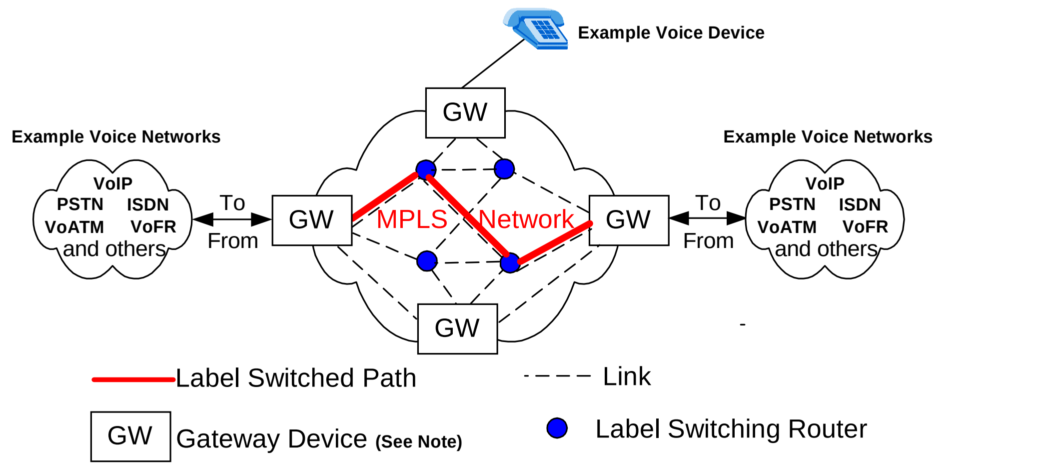
[5] ITU-T Recommendation G.729 — Coding of Speech at 8kbit/s using Conjugate Structure - Algebraic Code Excited Linear Predictive (CS-ACELP) Coding, March 1996

[6] B. Davie et al, “MPLS using LDP and ATM VC switching” RFC 3035

2 Reference Architecture

2.1 General

Figure 2-1 identifies the Reference Architecture for Voice over MPLS (VoMPLS). The MPLS network contains a number of Gateway (GW) devices, Label Switching Routers (LSR), and Label Switched Paths (LSP). An example LSP is shown as a solid line in the figure. Gateways may be directly connected to each other or indirectly connected through a number of LSRs.



Note: Gateway interworking functions with other networks or devices will not be specified in this implementation agreement

Figure 2-1. VoMPLS Reference Architecture

A simple architecture is all that is required in order to understand the application of this implementation agreement. It is not the intent of this agreement to specify the internal details of MPLS networks, the signaling required supporting VoMPLS, or the architecture or functions of gateways and routers. There are many different examples of how VoMPLS may be implemented and deployed in a network. The intent of the reference architecture is to support all possible deployments of VoMPLS.

The Gateway (GW) contains the functionality of a Label Edge Router (LER) as well as many other functions.

The Gateway device interfaces the MPLS network with:

- Other media (e.g., TDM, IP, ATM, etc.);
- Another MPLS network;
- Other networks (e.g., VoIP, PSTN, VoATM, etc.); and
- With access devices.

This architecture must be capable of supporting many different LSP bearer arrangements to convey voice payloads in an MPLS environment. For example:

- One arrangement may be an end-to-end LSP established between two voice devices existing within a single MPLS domain.
- A second arrangement may be an LSP that has been established to support only a portion of the voice connection between the end devices.

In the second case, multiple LSPs may need to be concatenated to form an end-to-end connection; or perhaps interworking between an LSP and another type of bearer may be required.

NOTE — This is a common occurrence in the current ISDN/PSTN environment where multiple service providers may be involved in carrying the call between the end devices.

An MPLS domain might exist between the entry and exit gateway nodes of the service provider network. LSPs are created between these network gateways to carry calls in a voice trunking arrangement.

2.2 Multiplexing voice calls onto MPLS Label Switched Paths (LSP)

Multiple voice calls may be transported over an LSP. Two types of VoMPLS subframes are defined, Primary and Control, and may be transmitted as required. Multiple primary subframes may be multiplexed within a single MPLS frame. The control subframes are not multiplexed and are sent separately; that is, only one control subframe at a time may be carried within an MPLS frame. Primary subframes and control subframes are not multiplexed together within a single MPLS frame.

A primary payload contains the traffic that is fundamental to the operation of a connection identified by a Channel Identifier (CID). It includes encoded voice and silence information descriptor(s). Primary payloads are variable length subframes.

Control subframes may be sent to support the primary payload (e.g., dialed digits for a primary payload of encoded voice) and other control functions. These payloads are differentiated from the primary payload by a Payload Type value in the subframe header. A range of payload type values is assigned to primary payload and control payloads. Control subframes are fixed length and most of them are sent with a triple redundant transmission with a fixed interval between them. The CID and payload type fields are common to both primary and control payload formats.

2.2.1 Primary Subframe

The MPLS frame structure allowing the multiplexing of primary subframes of Voice over MPLS calls is shown in Figure 2-2.

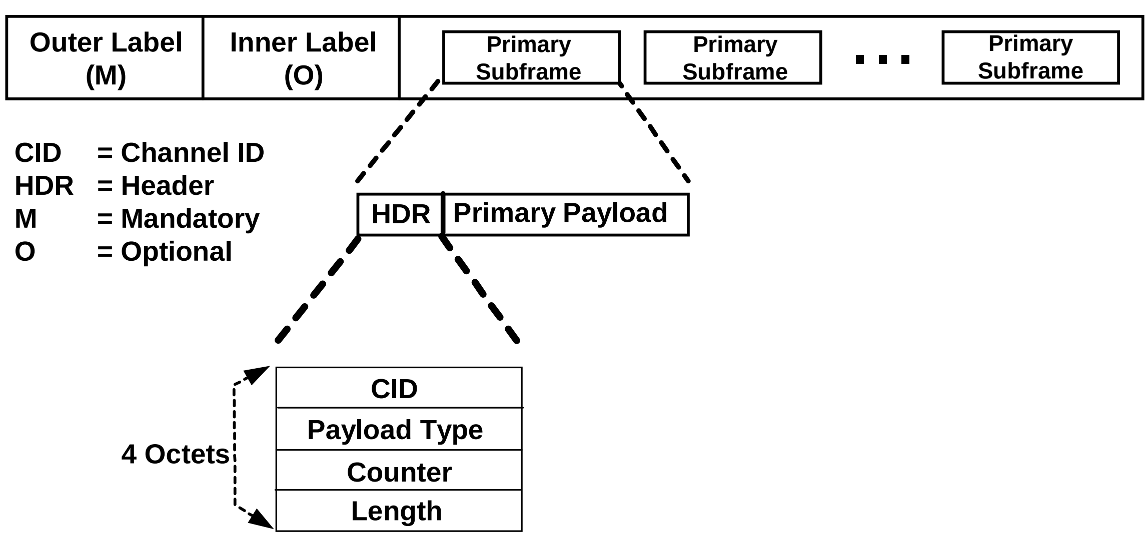


Figure 2-2. LSP Structure for Multiplexing Primary subframes of Voice Calls

A typical VoMPLS multiplexing structure consists of a mandatory outer label, zero or more inner labels, and one or more VoMPLS primary subframes consisting of a 4-octet header and variable length primary payload.

The Channel ID (CID) allows up to 248 VoMPLS calls to be multiplexed within a single LSP. At least one LSP must be created to convey VoMPLS calls; thus the use of an outer label is Mandatory. As an implementation option, additional inner LSPs may be created using stacked labels.

Figure 2-3 depicts an example VoMPLS primary frame structure of a single LSP that is used to convey from one to 248 VoMPLS channels. Note that a unique CID identifies each VoMPLS subframe but that the primary subframes may be transmitted in any order whenever information for a channel is available.

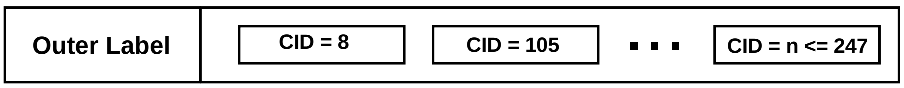


Figure 2-3. Single-LSP structure for multiplexing Primary Payloads of VoMPLS calls

In order to establish the single-LSP Voice over MPLS bearer structure depicted in Figure 2-3 the procedure is as follows:

1. A bi-directional LSP is created either by manual provisioning or by using an MPLS control protocol (e.g., CR-LDP, RSVP-TE).

2. As voice or audio connections arrive at the LER, a CID value is assigned to the connection (multiplexed) within the LSP. This is accomplished by either:

   a) A priori coordination of CID value usage. In this case each new call is assigned to an existing CID (i.e., there is no need for per call signaling);

   or

   b) An invocation of the signaling control protocol for CIDs to establish bi-directional channels that are used for the audio or voice connection. Note: identification or development of a signaling protocol for this purpose is not the subject of this implementation agreement.

Figure 2-4 depicts an example VoMPLS Primary frame structure based on label stacked inner LSPs. The outer label is the same while different inner labels are stacked to expand the multiplexing capability of the outer LSP.

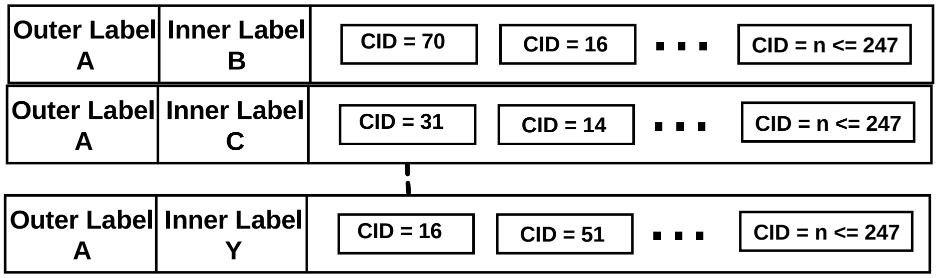


Figure 2-4. Stacked-LSP structure for multiplexing Primary payloads of VoMPLS calls

NOTE — A CID that is unique within each inner LSP identifies each VoMPLS subframe. That is, CID 16 in LSP-AB is a different channel than CID 16 in LSP-AY.

Both control and primary subframes may be transmitted in any order whenever information for a channel is available. This structure has the potential to convey up to 248 VoMPLS Channels multiplied by the number of inner LSPs.

In order to establish the stacked-LSP VoMPLS bearer structure depicted in Figure 2-4 the procedure is as follows:

1. A bi-directional LSP is created either by manual provisioning or by using an MPLS control protocol (e.g., CR-LDP, RSVP-TE). This LSP is termed the outer LSP.
2. As voice or audio connections arrive at the LER, an additional LSP may have to be created (multiplexed) within the outer LSP. This is accomplished by:
   a) Repeated invocations of the MPLS control protocol to establish bi-directional inner LSPs that are used for the voice or audio connection;

   or

   b) A-priori coordination of inner LSP label value usage. In this case each new call is assigned to an existing LSP (i.e. there is no need for per-call signaling).
3. As voice or audio connections arrive at the LER, a CID value is assigned to the connection (multiplexed) within the inner LSP. This is accomplished by:
   a) A-priori coordination of CID value usage. In this case each new call is assigned to an existing CID (i.e. there is no need for per-call signaling);

   or

   b) An invocation of a signaling control protocol for CIDs to establish bi-directional channels that are used for the voice or audio connection.

2.2.2 Control Subframe

The MPLS frame structure for control subframes of Voice over MPLS calls is shown in Figure 2-5.

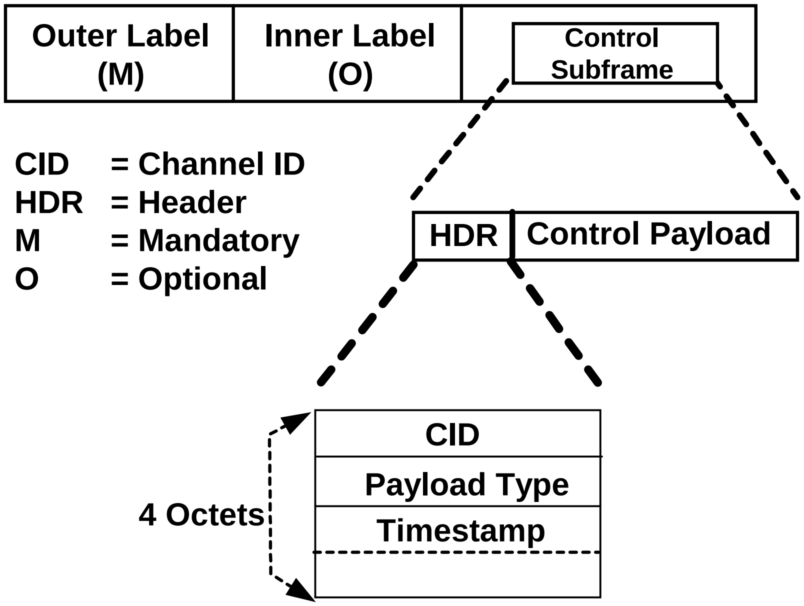


Figure 2-5. LSP Structure for Control Subframe in a VoMPLS call

3 Service Description

3.1 Primary Payloads

An MPLS frame containing VoMPLS primary payloads consists of the MPLS Label(s) followed by a sequence of primary subframes Each primary subframe consists of a header and a primary payload; each primary subframe may be associated with a different voice connection. A primary payload is either a sequence of encoded voice subframe(s) or a single Silence Insertion Descriptor subframe.

3.1.1 Encoded Voice

This service element conveys voice information supplied by the service user. The voice information is packaged according to the rules specified by voice transfer syntax. Various voice encoding algorithms that are supported in RFC 1890 and ITU-T Recommendation I.366.2 are supported in this Implementation agreement. Several transfer syntax definitions for voice compression schemes are described in the annexes of this Implementation Agreement.

Example voice encoding algorithms that are included in the Annexes are:

- G.711 – 64 Kbps (PCM);
- G.726 – 32 (ADPCM);
- G.723.1 – (MP-MLQ);
- G.729 – (CS-ACELP).

3.1.2 Silence Information Descriptor

Silence Information Descriptor (SID) subframes indicates the end of a talk-spurt and conveys comfort noise generation parameters. These SID indications support voice activity detection (VAD) and silence suppression schemes.

When VAD is utilized, a SID subframe may optionally be transmitted following the last encoded voice subframe of a talk-spurt. Reception of a SID subframe after a voice subframe may be interpreted as an explicit indication of end of talk-spurt. In addition, SID subframes may be transmitted at any time during the silence interval to update comfort noise generation parameters.

The SID payload is defined for PCM and ADPCM encoding. SID subframes should not be sent if VAD is not utilized.

The comfort noise analysis and synthesis as well as the VAD and discontinuous transmission algorithms are implementation specific.

3.2 Control Payload

The control payload consists of a single control subframe. The control subframe consists of a header and a control payload; the control subframe is associated with a specific voice connection.

3.2.1 Dialed Digits

This service element transparently conveys DTMF, or other dialed digits supplied by the service user. These digits may be sent during the voice call setup or following call establishment to transfer in-band tones.

Since some of the low bit-rate coding algorithms used may not properly pass the DTMF tones or other dialed digits, special capabilities must be employed to ensure the tones are properly conveyed.

3.2.2 Signaling Bits (Channel Associated Signaling)

This service element transparently conveys signaling bits supplied by the service user. These bits may indicate seizure and release of a connection, dial pulses, ringing, or other information in accordance with the signaling system in use over the transmission facility.

4 Additional Requirements

4.1 VoMPLS over ATM

When VoMPLS is operated over an ATM network, it shall follow RFC 3035 “MPLS using LDP and ATM switching”.

5 Frame Formats

5.1 General Format

The CID of a primary subframe or control subframe will identify the connection and serves as channel identification. As specified in Section 3, there are two types of protocol data units that are transported in an MPLS payload carrying VoMPLS:

- Primary payload with voice or audio information (e.g., encoded voice, Generic Silence Insertion Descriptor (SID), etc) and
- Control payload (e.g., signaling payload (dialed digits, channel associated signaling bits), etc).

5.2 Format of the Primary Subframe

The format of the primary subframe is shown in Figure 5-1. To maintain word (32 bits) alignment, the payload information must be a multiple of 4 octets. If the payload is not a multiple of 4 octets, up to 3 PAD octets are included to make it word aligned.

As specified in Section 3.1, a primary payload is either a sequence of encoded voice subframes or a single Silence Insertion Descriptor subframe.

The encoded voice subframe consists of one or more audio frames containing sample intervals or frames. The sample intervals or frames are placed sequentially in an encoded voice subframe. If the number of sample intervals or frames in a payload is more than one, the next interval or frame starts on the next octet after the previous interval or frame. PAD octets are only used in the last word of the payload if needed for word alignment.

NOTE – For a G.729 single interval (10 ms), the payload contains 2 octets of PAD (see Figure E-3). If two intervals are included in the payload, the second frame will start from octet 11. The total number of octets with two intervals in a payload is 20. No PAD octets are required (see Figure E-4 with M=2).

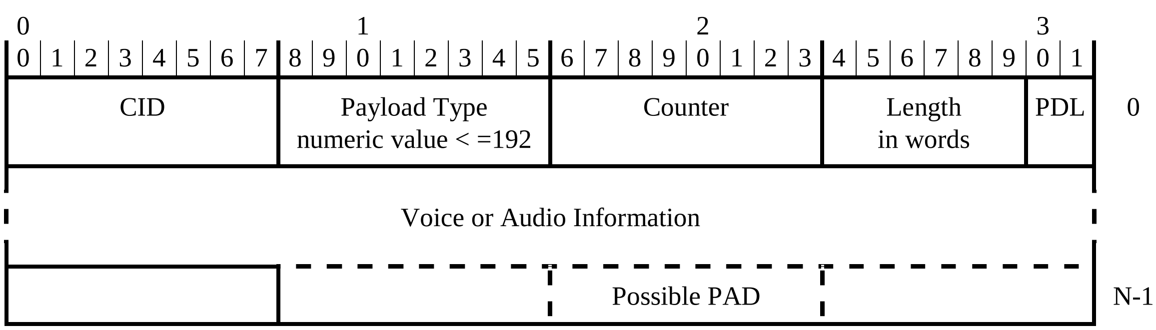


Figure 5-1. Format of the Primary payload

The fields in the header of primary payload frames are specified as follows:

a) Channel Identifier (CID):
This Channel Identifier indicates uniquely a voice or audio connection within the LSP of Voice over MPLS. The values "0" to "247" can be used to identify the VoMPLS user channels.

Table 5-1 – Coding of the CID Field

CID value	Use
0 to 247	Identification of VoMPLS user channels
248 249 250 to 255	Reserved for Layer Management peer-to-peer procedures Reserved for Signaling Reserved

b) Payload Type:
The payload type field indicates the payload type and encoding algorithm used for the voice or audio. The primary payload type field is coded according to Annex A. Negotiation of payload type for the connection is outside the scope of this Implementation Agreement.

Table 5-2 – Allocation of Payload Type values

Payload Type	Use
0 to 192	Allocated for Primary payloads (see Annex A)
193 to 223	Reserved
224 to 255	Allocated for Control payloads (see Table 5-3)

c) Counter:
The counter field provides a counter value at the first sample or frame in an encoded voice subframe. The initial value of the counter is derived from the initial timestamp for the connection (See section 7.3 for initialization and increment). After reaching the maximum unsigned count, the counter wraps around to zero.

d) Length:
The length indicates the number of voice/audio words (32 bits) in the voice frame including the PAD octets. It does not include the 4-octet header.

e) PAD Length (PDL):
The PDL field indicates the number of PAD octets in the last word (4 octets) of the primary payload.

5.3 Format of the Control Subframe

The format of the control payload frame is shown in Figure 5-2. In order to maintain word (32 bits) alignment, the control frame payload must be multiple of 4 octets. The length of the Control subframe is always inferred from its type.

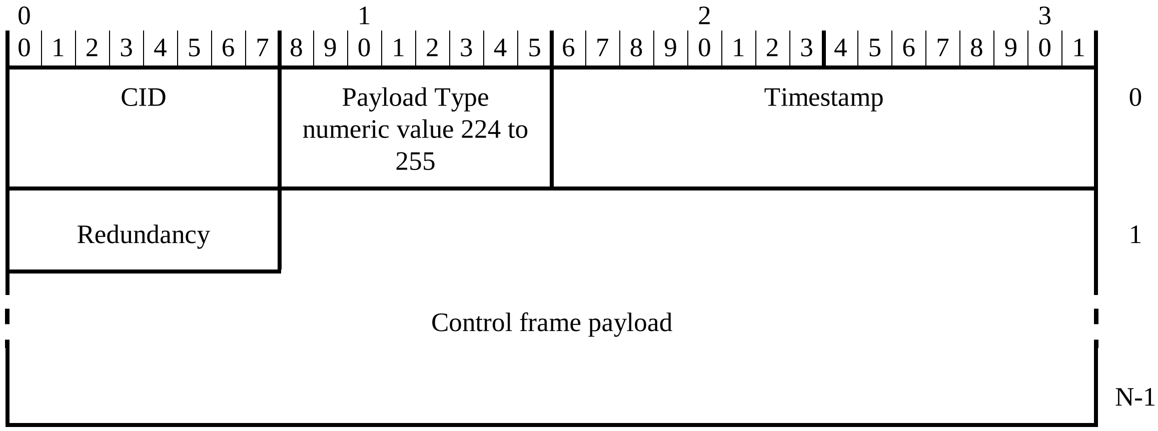


Figure 5-2. Format of the Control payload

The fields in the header of Control payload frames are specified as follows:

a) Channel Identifier (CID):
See section 5.2 item (a).

b) Payload Type:
The payload type field indicates the payload type of control payloads. The payload types for control frames are specified in Table 5-3.

Table 5-3 — Packet Types for Control Payloads

Payload Type	Description of the Control Payload	Reference
240	Dialed Digits	Annex G
241	Channel Associated Signaling	Annex H

c) Timestamp:
The timestamp reflects the sampling time of the control payload and it is coded in 125 μs units (8 kHz clock). It provides relative time. The initial value of the timestamp is random.

d) Redundancy:
The Redundancy field is set to values 0, 1 and 2 respectively for a packet's first, second, and third transmission under triple redundancy. Redundancy value 3 indicates no use of triple redundancy, whereby the payload is sent once.

6 Minimum Requirements for Conformance

This Implementation Agreement provides support for several optional transfer syntax definitions. Interoperability between VoMPLS devices is possible only when both devices provide support for one or more common transfer syntax definitions. An implementation is compliant with this agreement, if the following requirements are supported.

6.1 Frame Formats

1. Support the frame structure for “Single-LSP structure for multiplexing VoMPLS calls” as described in section 2 and figure 2-3.

2. Support for the “Stacked-LSP structure for multiplexing VoMPLS calls” as described in section 2 (Figure 2-4) is optional.

6.2 Primary payload types

1. Support of G.711-64 A-law and μ-law with a duration of 10 milliseconds as described in Annex B is mandatory.

2. Support of silence suppression is mandatory for G.711. The comfort noise analysis and synthesis as well as the VAD and discontinuous transmission algorithms are implementation specific. Support for the Generic Silence Insertion Descriptor is optional for transmitters and mandatory at the receivers.

3. Support for other primary payload transfer syntax definitions is optional.

6.3 Control payload types

1. Support for the signaled payload type (CAS) is mandatory at the receiver.

2. Support for other control payload types (e.g., dialed digit control payload type) is optional.

7 Procedures

7.1 General

One of the requirements of VoMPLS is that a stimulus occurring at the transmitter should be reproduced with the same interval between events at the receiver. Equivalently, the end-to-end delay of the information stream should be constant.

To keep the end-to-end delay constant, a receiver must have enough timing information to remove packet delay variation.

7.2 Audio encoding

This Implementation agreement identifies a set of encodings, each of which is comprised of a particular audio encoding and audio payload format for inclusion within an MPLS frame for VoMPLS. Some of those payload formats are specified in this Implementation Agreement, while others are specified in the revised RFC 1890 and ITU-T Recommendation I.366.2 (revised).

Annex A contains all payload types, characteristics of all audio encoding and default operating parameters. Applications should be prepared to handle other values. The default parameters allow a set of applications conforming to these guidelines to interoperate without additional negotiation or configuration.

The default interval of an encoded voice subframes should have duration of 10 ms, unless otherwise noted in Annex A.

NOTE — This packetization interval determines the minimum end-to-end delay. In VoMPLS the header overheads are low due to multiplexing of audio subframes, and a lower default value of 10 ms is chosen to minimize the end-to-end delay.

Encoding formats for ITU-T audio algorithms, including silence insertion, are defined in Annexes B through E.

7.2.1 Sample-based audio encoding

In sample-based encoding, each audio sample is represented by a fixed number of bits. An audio frame may contain a number of audio samples (e.g., for an interval of 5ms). The encoded voice subframe consists of one or more audio frames, each audio frame consisting of samples (e.g., from each 5 ms interval). The audio frames are placed sequentially in an encoded voice subframe.

The interval for an audio frame should be 5 ms. The number (M) of audio frames in an encoded voice subframe is configured or negotiated per channel at transmitter and receiver. It is typically, but not necessarily, the same in both directions. The default value of M is 2.

The number of samples included in an audio frame determines the duration of an audio frame. The packing of sample-based encoding producing less than one octet per sample is encoding specific.

The Counter field in the primary subframe header reflects the instant at which the first sample in the frame was sampled; that is, the oldest information in the frame.

7.2.2 Frame-based audio encoding

Frame-based encoding encodes a fixed length block into another block of compressed data, typically also of fixed length. For frame-based encoding, the sender may choose to combine several such audio frames into a single encoded voice subframe. All the data in a single encoded voice subframe shall be of the same encoding format, i.e., the same audio algorithm and bit rate. The receiver can tell the number of frames in the payload, by dividing the payload length by the audio frame size, which is defined as part of the encoding.

NOTE — This does not work when carrying frames of different sizes unless the frame sizes are relatively prime.

The default interval of an encoded voice subframe should have a duration of 10 ms or one frame; whichever is longer, unless otherwise noted in Annex A.

The frames are placed sequentially in the encoded voice subframe, so that the oldest frame occurs immediately after the primary subframe header.

The Counter field in the primary subframe header reflects the instant at which the first audio frame was sampled.

7.3 Timestamp and Counter

7.3.1 Timestamp

Timestamp field is included in the control subframe and reflects the sampling time of the control payload. It serves to counter packet delay variation and allows a receiver to reproduce accurately the relative timing of successive events that are separated by a short interval. Events that are separated by a long interval, e.g., many times the maximum packet delay variation, do not normally require precise timing.

The Timestamp field is 16 bits. The transmitter begins time stamping at an arbitrary value and increments by one every 125 µs (8 kHz clock). After reaching the maximum unsigned count, the time stamp wraps around to zero.

Having received two control frames that designate events, E1 and E2, with respective time stamps TS1 and TS2, a receiver should decide whether the interval between receptions of the packets is short enough to require precise timing of the events. If so, the receiver shall schedule their play-out times, PT1 and PT2, so that

PT2 – PT1 = TS2 - TS1.

NOTE — When performing subtraction, the result should be interpreted modulo 65536. This is important if the result of the subtraction is negative, i.e., if the time stamp of second event is less than the first.

7.3.2 Counter

The Counter field serves to counter packet delay variation and allows a receiver to reproduce accurately the relative successive events that are separated by a short interval. The Counter value is initialized to a value derived from the initial timestamp field (timestamp divided by 20) and incremented once every 2.5 ms. It allows 0.64 sec before wrapping.

NOTE — Since the primary purpose of the jitter buffer is to handle reordering, the capacity of the jitter buffer is typically 10 times the duration needed. Note that 2.5 ms is one half of the default 5 ms audio frame interval.

It is mandatory for the transmitter to increment the counter value from the preceding frame. The increment is an integer value equal to the duration of the preceding frame in milliseconds divided by 2.5. It is optional for a receiver to act on Counter value, and the algorithms that it may use are implementation specific. The Counter value of an audio frame shall correspond to the beginning of the sample or frame.

The Counter shall increment during periods of silence according to the Counter value of the last audio frame transmitted. This increment, when no frames are being transmitted, maintains the counter as relative time. The Counters shall not be reset at the beginning or end of a talk spurt.

NOTE — The reason for incrementing audio Counter through a silence period is to position the play out of the next talk spurt accurately with respect to the end of preceding talk spurt. This is a way to eliminate variation in the duration of silence between transmitter and receiver that might otherwise occur.

7.4 Triple redundancy

The common facility for control payloads requiring error correction is triple redundant transmission. Such control payloads are sent three times, with a fixed interval between transmissions.

The redundancy interval depends on the information stream. It is 5 ms for dialed digits and channel associated signaling bits.

Each copy of redundant control payloads contains the same content, except in the Redundancy field of the control subframe header. The three copies of a packet can be correlated because they all have the same time stamp.

The Redundancy field is set to values 0, 1 and 2 respectively for a packet's first, second, and third transmission under triple redundancy.

Redundancy value 3 indicates no use of triple redundancy, whereby some control payloads with the same format are sent singly, as specified in the corresponding Annex. These control payloads may occur periodically, but at a much longer interval. The receiver shall not expect three copies of them to be spaced at the redundancy interval.

NOTE — One use of this is for a long-term refresh of state information, such as the values of the CAS bits.

Annex A
Payload Types for Primary Payloads

(This Annex forms an integral part of this Implementation Agreement)

Payload Type	Description of Algorithm	Sample /Frame	Encoding Format Reference	Frame Time default (ms)	Sequence counter increment
0	G.711-64 (μ-law)	Sample	Annex B – Figure B-1	10	4
1	FED-STD 1016	Frame – 30 ms	RFC 1890 (Note 2)	30	12
2	G.726-32	Sample	Annex C - Figure C-1	10	4
4	G.723	Frame – 30 ms	RFC 1890 (Note 2)	30	12
5	DVI4-8000Hz	Sample	RFC 1890 (Note 2)	20	8
6	DVI4-16000Hz	Sample	RFC 1890 (Note 2)	20	8
7	LPC	Frame – 20 ms	RFC 1890 (Note 2)	20	8
8	G.711-64 (A-law)	Sample	Annex B – Figure B-1	10	4
9	G722-64	Sample	I.366.2 Annex C	20	8
11	L16	Sample	RFC 1890 (Note 2)	20	8
12	QCELP	Frame – 20 ms	RFC 1890 (Note 2)	20	8
15	G.728-16	Frame- 2.5ms	I.366.2 Annex G	20	8
16	DVI4-11025Hz	Sample	RFC 1890 (Note 2)	20	8
17	DVI4-22050Hz	Sample	RFC 1890 (Note 2)	20	8
18	G.729 or G.729A	Frame- 10ms	Annex E – Figure E-1	20	8
33	Generic SID		Annex F – Figure F-1		
35	G.711-56 (A-law)	Sample	I.366.2 Annex B	10	4
36	G.711-56 (μ-law)	Sample	I.366.2 Annex B	10	4
37	G.711-48 (A-law)	Sample	I.366.2 Annex B	10	4
38	G.711-48 (μ-law)	Sample	I.366.2 Annex B	10	4
39	G722-56	Sample	I.366.2 Annex C	20	8
40	G722-48	Sample	I.366.2 Annex C	20	8
41	G726-40	Sample	I.366.2 Annex E	10	4
42	G726-24	Sample	I.366.2 Annex E	10	4

Payload Type	Description of Algorithm	Sample /Frame	Encoding Format Reference	Frame Time default (ms)	Sequence counter increment
43	G726-16	Sample	I.366.2 Annex E	10	4
44	G727- (5,2), (5,3) and (5,4)	Sample	I.366.2 Annex F	20	8
45	G727- (4,2), (4,3) and (4,4)	Sample	I.366.2 Annex F	20	8
46	G727- (3,2) and (3,3)	Sample	I.366.2 Annex F	20	8
47	G727- (2,2)	Sample	I.366.2 Annex F	20	8
48	G.728-12.8	Frame- 2.5ms	I.366.2 Annex G	20	8
49	G.728-9.6	Frame- 2.5ms	I.366.2 Annex G	20	8
50	G.729D	Frame- 10ms	I.366.2 Annex H	20	8
51	G.729E	Frame- 10ms	I.366.2 Annex H	20	8
52	AMR 12.2	Frame- 20ms	I.366.2 Annex Q	20	8
53	AMR 12.2 (errored)	Frame- 20ms	I.366.2 Annex Q	20	8
54	AMR 10.2	Frame- 20ms	I.366.2 Annex Q	20	8
55	AMR 10.2 (errored)	Frame- 20ms	I.366.2 Annex Q	20	8
56	AMR 7.95	Frame- 20ms	I.366.2 Annex Q	20	8
57	AMR 7.95 (errored)	Frame- 20ms	I.366.2 Annex Q	20	8
58	AMR 7.4	Frame- 20ms	I.366.2 Annex Q	20	8
59	AMR 7.4 (errored)	Frame- 20ms	I.366.2 Annex Q	20	8
60	AMR 6.7	Frame- 20ms	I.366.2 Annex Q	20	8
61	AMR 6.7 (errored)	Frame- 20ms	I.366.2 Annex Q	20	8
62	AMR 5.9	Frame- 20ms	I.366.2 Annex Q	20	8
63	AMR 5.9 (errored)	Frame- 20ms	I.366.2 Annex Q	20	8
64	AMR 5.15	Frame- 20ms	I.366.2 Annex Q	20	8
65	AMR 5.15 (errored)	Frame- 20ms	I.366.2 Annex Q	20	8
66	AMR 4.75	Frame- 20ms	I.366.2 Annex Q	20	8
67	AMR 4.75 (errored)	Frame- 20ms	I.366.2 Annex Q	20	8
68	AMR SID		I.366.2 Annex Q	160	64
69	Unspecified (note 4)	Sample		10	4

Payload Type	Description of Algorithm	Sample /Frame	Encoding Format Reference	Frame Time default (ms)	Sequence counter increment
Other values	Reserved				

Note 1 – The payload type codepoints from 0 to 18 are from Table 4 "payload types (PT) for audio encodings" in Draft RFC draft-ietf-avt-profile-new-09.txt.

Note 2 – For encoding formats see Draft RFC draft-ietf-avt-profile-new-09.txt "RTP Profile for Audio And Video Conference with Minimal Control (revision of RFC 1890).

Note 3 – The AMR xx (errored) code points provide Frame Quality Classification (FQC) (see AMR Speech Codec Frame Structure - 3G TS 26.101 version 3.0.0 Release 1999) indication. FQC indictor is not in the AMR payload.

Note 4 – The unspecified code point is used to transport the information transparently (e.g., transport information between two ISDN subnets).

Annex B
Encoding format for audio algorithm G.711

(This Annex forms an integral part of this Implementation Agreement)

B.1 General

G.711 Pulse Code Modulation (PCM) is a coder that produces one 8-bit value every 125 μs representing the sign and amplitude of an audio sample. Two encoding laws are recommended, referred to as A-law and μ-law.

Encoded values are represented with the polarity (sign) bit as the most significant bit (see Tables 1/G.711 and 2/G.711). Bit numbering convention adopted here is based on RFC 791 (1 to 8 of G.711 is mapped to 0 to 7). Recommendation G.711 does not define an intrinsic SID and may be used with generic SID of Annex F.

B.2 Audio frame format of sampled speech

For G.711 audio encoding, sampled-based encoding is used with a fixed interval of 5 ms. In each interval; G.711 audio samples are accumulated to yield a sequence of 40 encoded samples. These are concatenated in chronological order, with the earliest positioned at the most significant bit of the first octet.

Formats for G.711 64 kbit/s is shown in Figures B-1. They are the same for A-law, and μ-law PCM.

0	1	2	3	
0 1 2 3 4 5 6 7	8 9 0 1 2 3 4 5	6 7 8 9 0 1 2 3	4 5 6 7 8 9 0 1	
PCM 1	PCM 2	PCM 3	PCM 4	0
PCM 5	PCM 6	PCM 7	PCM 8	1
PCM 9	PCM 10	PCM 11	PCM 12	2
PCM 29	PCM 30	PCM 31	PCM 32	7
PCM 33	PCM 34	PCM 35	PCM 36	8
PCM 37	PCM 38	PCM 39	PCM 40	9

Figure B-1. G.711-64 Audio frame format

A G.711 encoded voice subframe consists of one or more audio frames, each audio frame consisting of 40 samples of one octet each (5 ms per interval).

B.3 Transfer characteristics

Encoding interval of audio frame: 5 ms

Number of audio frames in encoded voice subframe: M = 1 to 6

Support of M = 2 is required.

A range of 1 to 6 can optionally be supported.

Packetization time: **5***M ms

Payload Type	Description of algorithm	Encoding format reference	Bit Rate (kbit/s)	Voice Transfer Structure (Octets)
8	G.711-64 (A-law)	Figure B-1	64	40 * M
0	G.711-64 (μ-law)	Figure B-1	64	40 * M

B.4 Examples of G.711-64 subframe

0 1 2 3

0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1

CID value	**Payload type = 8**	**Counter value**	**Length = 20**	**PDL = 0**	0
PCM 1	PCM 2	PCM 3	PCM 4		1
PCM 5	PCM 6	PCM 7	PCM 8		2
PCM 9	PCM 10	PCM 11	PCM 12		3
PCM 69	PCM 70	PCM 71	PCM 72		18
PCM 73	PCM 74	PCM 75	PCM 76		19
PCM 77	PCM 78	PCM 79	PCM 80		20

Figure B-2. G.711-64 A-law primary payload with M=2

0 0 1 2 3 4 5 6 7	1 8 9 0 1 2 3 4 5	2 6 7 8 9 0 1 2 3	3 4 5 6 7 8 9 0 1	
CID value	**Payload type = 0**	**Counter value**	**Length = 20** | **PDL = 0**	0
PCM 1	PCM 2	PCM 3	PCM 4	1
PCM 5	PCM 6	PCM 7	PCM 8	2
PCM 9	PCM 10	PCM 11	PCM 12	3
PCM 69	PCM 70	PCM 71	PCM 72	18
PCM 73	PCM 74	PCM 75	PCM 76	19
PCM 77	PCM 78	PCM 79	PCM 80	20

Figure B-3. G.711-64 μ-law primary payload with M=2

Annex C
Encoding format for audio algorithm G.726-32

(This Annex forms an integral part of this Implementation Agreement)

C.1 General

G.726-32 Adaptive Pulse Code Modulation (ADPCM) supports bit rates of 32 kbit/s. The encoding produces 4 bits, every 125 µs.

Encoded values are represented with the sign bit as the most significant bit (see Tables 7/G.726 through 10/G.726).

Recommendation G.726 does not define an intrinsic SID and may be used with the generic SID of Annex F. If this is the case, the audio coder and decoder shall be reset synchronously at the beginning of each talk spurt, as described in Annex F.3.

C.2 Audio frame format of sampled speech

The audio frame format requires that G.726 outputs be accumulated over an interval of 5 ms to yield a sequence of 40 encoded values. These are concatenated in chronological order, with the earliest positioned at the most significant bit of the first octet.

Formats for the coding rates of 32 kbit/s are shown in Figure C-1.

0 1 2 3	4 5 6 7	8 9 0 1	2 3 4 5	6 7 8 9	0 1 2 3	4 5 6 7	8 9 0 1	
Adpcm 1	Adpcm 2	Adpcm 3	Adpcm 4	Adpcm 5	Adpcm 6	Adpcm 7	Adpcm 8	0
Adpcm 9	Adpcm 10	Adpcm 11	Adpcm 12	Adpcm 13	Adpcm 14	Adpcm 15	Adpcm 16	1
Adpcm 17	Adpcm 18	Adpcm 19	Adpcm 20	Adpcm 21	Adpcm 22	Adpcm 23	Adpcm 24	2
Adpcm 25	Adpcm 26	Adpcm 27	Adpcm 28	Adpcm 29	Adpcm 30	Adpcm 31	Adpcm 32	3
Adpcm 33	Adpcm 34	Adpcm 35	Adpcm 36	Adpcm 37	Adpcm 38	Adpcm 39	Adpcm 40	4

Figure C-1. G.726-32 Audio frame format

A G.726-32 encoded voice subframe consists of one or more audio frames; each audio frame consisting of 40 samples of 4 bit each (5 ms per interval, packed in 20 octets)

C.3 Transfer characteristics

Encoding interval of audio frame: 5 ms

Number of audio frames in encoded voice subframe: M = 1 to 6

A range of 1 to 6 can optionally be supported.
Default M = 2.

Packetization time: **5***M ms

Payload Type	Description of algorithm	Encoding format reference	Compression Bit Rate (kbit/s)	Voice Transfer Structure (Octets)
2	G.726-32 ADPCM	Figure C-1	32	20 * M

C.4 Example of G.726-32 subframe

0 1 2 3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1

CID value		**Payload type = 2**		**Counter value**		**Length = 10**	**PDL = 0**	0
Adpcm 1	Adpcm 2	Adpcm 3	Adpcm 4	Adpcm 5	Adpcm 6	Adpcm 7	Adpcm 8	1
Adpcm 9	Adpcm 10	Adpcm 11	Adpcm 12	Adpcm 13	Adpcm 14	Adpcm 15	Adpcm 16	2
Adpcm 65	Adpcm 66	Adpcm 67	Adpcm 68	Adpcm 69	Adpcm 70	Adpcm 71	Adpcm 72	9
Adpcm 73	Adpcm 74	Adpcm 75	Adpcm 76	Adpcm 77	Adpcm 78	Adpcm 79	Adpcm 80	10

Figure C-2. G.726-32 Primary payload with M=2

Annex D
Encoding format for audio algorithm G.723

(This Annex forms an integral part of this Implementation Agreement)

D.1 General

G.723 is specified in ITU Recommendation G.723.1, "Dual-rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s". The G.723.1 operates at either 5.3 or 6.3 kbit/s. Both rates are a mandatory part of the encoder and decoder. The algorithm has a floating-point specification in G.723.1 Annex B, a silence compression algorithm in G.723.1 Annex A and an encoded signal bit-error sensitivity specification in G.723.1 Annex C.

Every 30 ms, G.723.1 emits either 160 or 192 bits, respectively, that characterize a voice sample with an additional delay of 7.5 ms due to look-ahead. A G.723.1 frame can be one of three sizes: 24 octets (6.3 kb/s frame), 20 octets (5.3 kb/s frame), or 4 octets. The 4-octet frames are called SID frames (Silence Insertion Descriptor) and are used to specify comfort noise parameters. There is no restriction on how these three frames are intermixed. The least significant two bits of the first octet in the frame determine the frame size and codec type. It is possible to switch between the two rates at any 30 ms boundary.

D.2 Audio frame format

The bits of a G.723.1 frame are formatted as shown in Figures D.1 and D.2 (see Tables 5/G.723.1 and 6/G.723.1). Within the fields of a data unit, later octets are more significant. This is based on H.324 bit order assignment and is the reverse of the RFC 791 convention. With each octet, the bits are with the MSB on the left and the LSB on the right.

<table>
<tr><th>0</th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th><th>6</th><th>7</th><th></th></tr>
<tr><td colspan="6">LPC_B5...B0</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td colspan="8">LPC_B13...B6</td><td>1</td></tr>
<tr><td colspan="8">LPC_B21...B14</td><td>2</td></tr>
<tr><td colspan="6">ACL0_B5...B0</td><td colspan="2">LPC_B23...B22</td><td>3</td></tr>
<tr><td colspan="5">ACL2_B4...B0</td><td colspan="2">ACL1_B1...B0</td><td>ACL0_B6</td><td>4</td></tr>
<tr><td colspan="4">GAIN0_B3...B0</td><td colspan="2">ACL3_B1...B0</td><td colspan="2">ACL2_B6...B5</td><td>5</td></tr>
<tr><td colspan="8">GAIN0_B11...B4</td><td>6</td></tr>
<tr><td colspan="8">GAIN1_B7...B0</td><td>7</td></tr>
<tr><td colspan="4">GAIN2_B3...B0</td><td colspan="4">GAIN1_B11...B8</td><td>8</td></tr>
<tr><td colspan="8">GAIN2_B11...B4</td><td>9</td></tr>
<tr><td colspan="8">GAIN3_B7...B0</td><td>10</td></tr>
<tr><td>GRID3</td><td>GRID2</td><td>GRID1</td><td>GRID0</td><td colspan="4">GAIN3_B11...B8</td><td>11</td></tr>
<tr><td colspan="7">MSBPOS_B6...B0</td><td>0</td><td>12</td></tr>
<tr><td colspan="2">POS0_B1...B0</td><td colspan="6">MSBPOS_B12...B7</td><td>13</td></tr>
<tr><td colspan="8">POS0_B9...B2</td><td>14</td></tr>
<tr><td colspan="2">POS1_B1...B0</td><td colspan="6">POS0_B15...B10</td><td>15</td></tr>
<tr><td colspan="8">POS1_B9...B2</td><td>16</td></tr>
<tr><td colspan="4">POS2_B3...B0</td><td colspan="4">POS1_B13...B10</td><td>17</td></tr>
<tr><td colspan="8">POS2_B11...B4</td><td>18</td></tr>
<tr><td colspan="4">POS3_B3...B0</td><td colspan="4">POS2_B15...B12</td><td>19</td></tr>
<tr><td colspan="8">POS3_B11...B4</td><td>20</td></tr>
<tr><td colspan="6">PSIG0_B5...B0</td><td colspan="2">POS3_B13...B12</td><td>21</td></tr>
<tr><td colspan="3">PSIG2_B2...B0</td><td colspan="5">PSIG1_B4...B0</td><td>22</td></tr>
<tr><td colspan="5">PSIG3_B4...B0</td><td colspan="3">PSIG2_B5...B3</td><td>23</td></tr>
</table>

Figure D-1. G.723.1-6.3 Audio frame format

<table>
<tr><th>0</th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th><th>6</th><th>7</th><th></th></tr>
<tr><td colspan="6">LPC_B5...B0</td><td>0</td><td>1</td><td>0</td></tr>
<tr><td colspan="8">LPC_B13...B6</td><td>1</td></tr>
<tr><td colspan="8">LPC_B21...B14</td><td>2</td></tr>
<tr><td colspan="6">ACL0_B5...B0</td><td colspan="2">LPC_B23...B22</td><td>3</td></tr>
<tr><td colspan="5">ACL2_B4...B0</td><td colspan="2">ACL1_B1...B0</td><td>ACL0_B6</td><td>4</td></tr>
<tr><td colspan="4">GAIN0_B3...B0</td><td colspan="2">ACL3_B1...B0</td><td colspan="2">ACL2_B6...B5</td><td>5</td></tr>
<tr><td colspan="8">GAIN0_B11...B4</td><td>6</td></tr>
<tr><td colspan="8">GAIN1_B7...B0</td><td>7</td></tr>
<tr><td colspan="4">GAIN2_B3...B0</td><td colspan="4">GAIN1_B11...B8</td><td>8</td></tr>
<tr><td colspan="8">GAIN2_B11...B4</td><td>9</td></tr>
<tr><td colspan="8">GAIN3_B7...B0</td><td>10</td></tr>
<tr><td>GRID3</td><td>GRID2</td><td>GRID1</td><td>GRID0</td><td colspan="4">GAIN3_B11...B8</td><td>11</td></tr>
<tr><td colspan="8">POS0_B7...B0</td><td>12</td></tr>
<tr><td colspan="4">POS1_B3...B0</td><td colspan="4">POS0_B11...B8</td><td>13</td></tr>
<tr><td colspan="8">POS1_B11...B4</td><td>14</td></tr>
<tr><td colspan="8">POS2_B7...B0</td><td>15</td></tr>
<tr><td colspan="4">POS3_B3...B0</td><td colspan="4">POS2_B11...B8</td><td>16</td></tr>
<tr><td colspan="8">POS3_B11...B4</td><td>17</td></tr>
<tr><td colspan="4">PSIG1_B3...B0</td><td colspan="4">PSIG0_B3...B0</td><td>18</td></tr>
<tr><td colspan="4">PSIG3_B3...B0</td><td colspan="4">PSIG2_B3...B0</td><td>19</td></tr>
</table>

Figure D-2. G.723.1-5.3 Audio frame format

A G.723.1 encoded voice subframe consistes of a single audio frame, the audio frame (30 ms per frame) is packs into 20 or 24 octets.

D.3 Silence Insertion Descriptor (SID)

Annex A/G.723.1 defines a voice activity detector and comfort noise generator for use with G.723.1. It classifies each 30 ms sample as either active voice or background noise.

Active voice is encoded according to Figures D-1 and D-2. Background noise is encoded as a Silence Insertion Descriptor according to Figure D-3 (see Table A.1/G.723.1). SIDs are sent only intermittently, when an appreciable change is detected in the nature of the background noise.

<table>
<tr><th>0</th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th><th>6</th><th>7</th><th></th></tr>
<tr><td colspan="6">LPC_B5...B0</td><td>1</td><td>0</td><td>0</td></tr>
<tr><td colspan="8">LPC_B13...B6</td><td>1</td></tr>
<tr><td colspan="8">LPC_B21...B14</td><td>2</td></tr>
<tr><td colspan="6">GAIN_B5...B0</td><td colspan="2">LPC_B23...B22</td><td>3</td></tr>
</table>

Figure D-3. G.723.1 SID packet format

NOTE — A single Silence Insertion Descriptor subframe is carried in a Primary subframe.

D.4 Transfer characteristics

Encoding interval of audio frame: 30 ms
Number of audio frames in encoded voice subframe: M = 1
Packetization time: **30** ms

Payload Type	Description of algorithm	Encoding format reference	Compression Bit Rate (kbit/s)	Frame Size (Octets)
4	G.723.1	Figure D-1	6.3	24
		Figure D-2	5.3	20

D.5 Examples of G.723 subframes

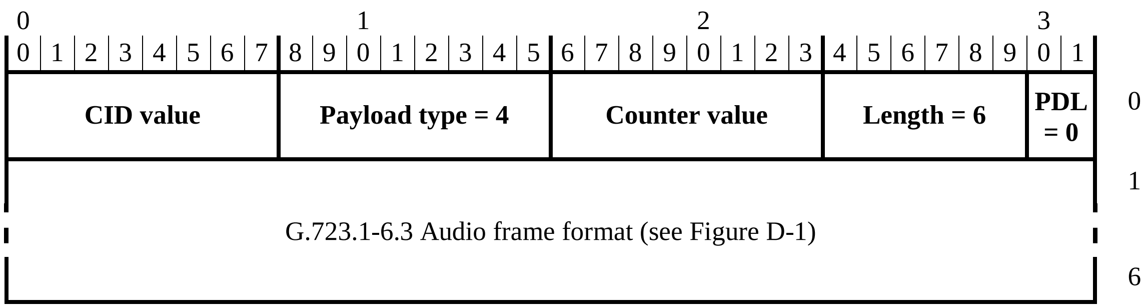


Figure D-4. G.723.1-6.3 subframe format

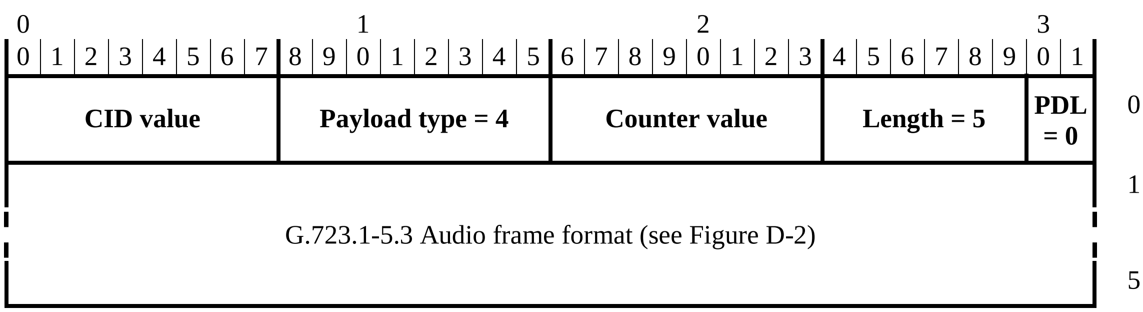


Figure D-5. G.723.1-5.3 subframe format

0 1 2 3 4 5 6 7	8 9 0 1 2 3 4 5	6 7 8 9 0 1 2 3	4 5 6 7 8 9	0 1	
CID value	**Payload type = 4**	**Counter value**	**Length = 1**	**PDL = 0**	0
G.723.1 SID packet format (see Figure D-3)					1

Figure D-6. G.723.1 SID subframe format

Annex E
Encoding format for audio algorithm G.729

(This Annex forms an integral part of this Implementation Agreement)

E.1 General

The basic algorithm of G.729 runs at 8 kbit/s. Every 10 ms it emits 80 bits that encode a voice frame. Encoded values are represented in this specification according to the conventions of RFC 791, whereby earlier octets and left most bits are more significant.

G.729 Annex A defines a reduced complexity coder that is interoperable with basic G.729. The format of the encoded values is the same for G.729, and G.729 Annex A. Any combination of G.729, and G.729 Annex A transmitter and receiver can be used together.

G.729 Annex B defines a voice activity detector and comfort noise generator for use with G.729 or G.729 Annex A. It classifies each 10 ms sample as either active voice or background noise.

E.2 G.729 audio frame format

The bits of a G.729 frame are formatted as shown in Figure E.1 (see Table 8/G.729). Within the fields of a data unit, bit and octet significance follows the RFC 791 convention adopted here.

<table>
<tr><th>0</th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th><th>6</th><th>7</th><th></th></tr>
<tr><td>L0</td><td colspan="7">L1</td><td>0</td></tr>
<tr><td colspan="5">L2</td><td colspan="3">L3</td><td>1</td></tr>
<tr><td colspan="2"></td><td colspan="6">P1</td><td>2</td></tr>
<tr><td colspan="2"></td><td>P0</td><td colspan="5"></td><td>3</td></tr>
<tr><td colspan="8">C1</td><td>4</td></tr>
<tr><td colspan="4">S1</td><td colspan="3">GA1</td><td></td><td>5</td></tr>
<tr><td colspan="3">GB1</td><td colspan="5">P2</td><td>6</td></tr>
<tr><td colspan="8">C2</td><td>7</td></tr>
<tr><td colspan="5"></td><td colspan="3">S2</td><td>8</td></tr>
<tr><td></td><td colspan="3">GA2</td><td colspan="4">GB2</td><td>9</td></tr>
</table>

Figure E-1. G.729-8 Audio frame format

A G.729 encoded voice subframe consists of one or more audio frames, each audio frame is packetized into 10 octets.

E.3 Silence Insertion Descriptor (SID)

Active voice is encoded according to Figure E.1. Background noise is encoded as a Silence Insertion Descriptor according to Figure E.2 (see Table B.2/G.729). SID is sent only intermittently, when an appreciable change is detected in the nature of the background noise.

0	1	2	3	4	5	6	7	
SPI	First stage vector							1
Second stage vector		Gain (energy)					RES	2

SPI Switched predictor index of LSF quantizer

RES Reserved (set to zero)

Figure E-2. G.729 Silence Insertion Descriptor frame format

E.4 Transfer characteristics

Encoding interval of audio frame: 10 ms

Number of audio frames in encoded voice subframe: M = 1 to 6

A range of 1 to 6 can optionally be supported.
Default M = 2.

Packetization time: M*10 ms

Voice Coding Type	Description of algorithm	Encoding format reference	Compression Bit Rate (kbit/s)	Frame Size (Octets)
18	G.729 or G.729A (CS-ACELP)	Figure E-1	8	M*10

E.5 Example of G.729 or G.729A subframes

0 1 2 3 4 5 6 7	8 9 0 1 2 3 4 5	6 7 8 9 0 1 2 3	4 5 6 7 8 9	0 1	
CID value	**Payload type = 18**	**Counter value**	**Length = 3**	**PDL = 2**	0
G.729-8 Audio frame format (see Figure E-1)					1
					2
		PAD			3

Figure E-3. G.729 or G.729A primary subframe format with M=1

0 1 2 3 4 5 6 7	8 9 0 1 2 3 4 5	6 7 8 9 0 1 2 3	4 5 6 7 8 9	0 1	
CID value	**Payload type = 18**	**Counter value**	**Length = 5**	**PDL = 0**	0
G.729-8 Audio frame 1 format (see Figure E-1)					1
					2
		G.729-8 Audio frame 2 format (see Figure E-1)			3
					4
					5

Figure E-4. G.729 or G.729A primary subframe format with M=2

0 1 2 3 4 5 6 7	8 9 0 1 2 3 4 5	6 7 8 9 0 1 2 3	4 5 6 7 8 9	0 1	
CID value	**Payload type = 18**	**Counter value**	**Length = 1**	**PDL = 2**	0
G.729 SID frame format (see Figure E-2)		PAD			1

Figure E-5. G.729 Silence Insertion Descriptor subframe format

Annex F
Encoding format for generic Silence Insertion Descriptor

(This Annex forms an integral part of this Implementation Agreement)

F.1 General

ITU-T Recommendations G.711, G.722, G.726, G.727 and G.728 do not contain provisions for voice activity detection, discontinuous transmission, and comfort noise generation tailored to the specific algorithm. Such procedures may be added in a generic way. This annex provides format for communication of comfort noise parameters. The comfort noise analysis and synthesis as well as the Voice Activity Detection (VAD) and Discontinuous Transmission algorithms are unspecified and left implementation specific.

The payload format is based on Appendix II of ITU-T Recommendation G.711. The comfort noise payload consists of a single octet description of the noise level.

F.2 Silence Insertion Descriptor subframe format

The Generic Silence Insertion Descriptor subframe format is shown in Figure F-1.

<table>
<tr><td>0</td><td>1 2 3 4 5 6 7</td><td>1
8 9 0 1 2 3 4 5</td><td>2
6 7 8 9 0 1 2 3</td><td>4 5 6 7 8 9</td><td>3
0 1</td><td></td></tr>
<tr><td colspan="2">CID value</td><td>Payload type = 33</td><td>Counter value</td><td>Length = 1</td><td>PDL =3</td><td>0</td></tr>
<tr><td>0</td><td>Noise Level</td><td colspan="4">PAD</td><td>1</td></tr>
</table>

Figure F-1. Generic Silence Insertion Descriptor subframe format

The Noise Level field is coded according to Table F-1. It represents the total noise power level that the transmitter wishes to convey to the receiver. Other noise characteristics, such as the spectral distribution, are not specified.

Table F-1. Noise level codes

Noise Level	Meaning
0-29	Reserved
30	–30 dBm0
31	–31 dBm0
. . .	. . .
77	–77 dBm0
78	–78 dBm0
79-126	Reserved
127	Idle Code (Note 2)

Note 1 – Table F-1 provides codepoints for measured noise levels, and the full or partial use of these codepoints is up to the implementation.

Note 2 – The receiver will use default noise level.

F.3 Procedures

The job of voice activity detection is to determine between active and inactive voice segments in the input signal. During inactive voice segments, the role of the noise generation algorithm is to sufficiently describe the ambient noise. A generic SID packet shall be sent immediately after the last active voice packet of a talk spurt. It marks the beginning of silence and alerts the receiver to expect an absence of active voice packets. The SID may be sent periodically during a silence period, or only when there is a significant change in the background noise characteristics.

The generic Silence Insertion Descriptor subframe may be multiplexed with primary payload frames in an MPLS payload.

The noise generation algorithm at the receiver uses the information in the SID to update its noise generation model and then produce an appropriate amount of comfort noise.

Since other characteristics of the noise, aside from its level, are not specified, they are to be chosen by the receiver. If a receiver is not capable of generating the total power level specified, it may generate a different level or may apply the idle code, i.e., default noise level. Otherwise, the specified level should be considered a guideline.

If the first active voice packet following a generic SID packet selects an adaptive audio algorithm – such as G.726 or G.727 – encoding and decoding of that packet shall be performed starting from an audio coder state that has been reset to its specified initial values.

NOTE – This eliminates glitches that could otherwise occur if the transmitter's state were to diverge during the silence, when active voice packets are not being sent and the receiver's state is not being updated. Resetting both encoder and decoder in this way maintains a synchronized state and initiates a fresh adaptation for each talk spurt, unbiased by the previous one.

Annex G
Packet Format and Procedures for Dialed Digits

(This Annex forms an integral part of this Implementation Agreement)

G.1 General

The dialed digits packet format can be used to transport Dual-Tone Multi-Frequency (DTMF) signals across an MPLS connection for reproduction at the other side. Recommendation Q.23 defines the frequency coding of DTMF. Payload carrying DTMF digits will be identified using payload type of 240.

Dialed digit packets constitute a separate, secondary information stream that avoids dependence on the audio encoding profile in effect. Some low bit rate audio encoding, such as G.723.1 do not convey multifrequency tones with acceptable fidelity. Other audio encodings that have higher fidelity may not require the support of dialed digit packets but can still find savings in bandwidth by the use of the dialed digits procedure.

The transmission of dialed digit packets is optional.

Dialed digits may be used in the middle of a call to convey user commands to a device at the far end of a connection, such as an automatic voice message recording system.

Dialed digit and audio encoding packets may occur at the same time. They are independent streams and may experience differential delay in reconstruction and play-out. In general, dialed digit signals are generated to be recognized by machines; during the intervals of operation other audio in the same direction is ignored. For the most reliable operation, transmitters should stop sending audio packets while detecting and sending dialed digits. However, specification of the behavior of transmitting users is beyond the scope of this Implementation Agreement. For this reason, receivers should discard any audio while playing out dialed digits instead of striving to merge the two streams.

G.2 Control subframe format

Dialed digit control subframe is sent with triple redundant transmission. The control subframes are sent three times, with a fixed interval between transmissions (see § 7.4.).

The format of dialed digit packets is shown in Figure G-1.

<table>
<tr><td colspan="8">0</td><td colspan="8">1</td><td colspan="8">2</td><td colspan="8">3</td><td></td></tr>
<tr><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>0</td><td>1</td><td></td></tr>
<tr><td colspan="8">CID value</td><td colspan="8">Payload type = 240</td><td colspan="16">Timestamp</td><td>0</td></tr>
<tr><td colspan="8">Redundancy</td><td colspan="8">Signal Level</td><td colspan="3">Digit Type</td><td colspan="5">Digit Code</td><td colspan="8">Reserved</td><td>1</td></tr>
</table>

Figure G-1. Dialed digits control subframe format

Signal Level: The Signal Level field is coded with a binary value from 0 to 31, signifying a total power level of 0 to –31 dBm0. Levels of –31 dBm0 and below are indicated by the value 31, and levels of 0 dBm0 and above are indicated by the value 0. All other values in the Signal Level field are reserved.

Digit Type:
The Digit Type field is coded according to Table G-1.

Table G-1. Dialed digits type codes

Digits Type	Meaning
000	Reserved
001	DTMF
010 to 111	Reserved

Digit Code:
The Digit Code field is coded according to Table G-2 for DTMF.

Table G-2. DTMF dialed digit codes

Digit Code	Meaning
00000	0
00001	1
00010	2
00011	3
00100	4
00101	5
00110	6
00111	7
01000	8
01001	9
01010	*
01011	#
01100	A
01101	B
01110	C
01111	D
10000 to 11110	Reserved
11111	Tone-off

G.3 Transmitter procedures

When a transmitter wishes to convey to the receiver the beginning of a dialed digit, the dialed digits control subframe shall be sent with triple redundancy at intervals of 5 ms.

If a tone persists, every 500 ms thereafter the dialed digits frame shall be sent to refresh the play-out (with the Redundancy field coded as value 3).

If a new event is conveyed before the triple redundancy of a previous event has completed, the transmitter shall stop sending control subframes for the previous event, in order to avoid the interleaving of two different time stamps.

A user transmitting dialed digits should ensure that no more than 20 ms of DTMF tones are allowed to pass through the encoded audio path, so that differential delays between the two streams do not cause false double signals at the far end receiver.

If a transmitter detects multifrequency tones but is not capable of determining their level, it shall set the Signal Level field to a preset value.

G.4 Receiver procedures

A user receiving dialed digits is expected to regenerate dialed digit signals according to the parameters conveyed to the best of its ability. At most one signal shall be played at any given time. Transitions are explicitly indicated, and a new dialed digit signal implicitly turns off the old one. Transitions to silence are explicit and triply redundant, like all others.

When regenerated, the two frequencies of a tone pair and their relative levels should be within the tolerances for the local environment. The total power level should be as indicated.

In order to make full use of triple redundancy without introducing extra delay variation, the receiver should wait before indicating dialed digits until such time as it should have received all three copies of a transition. Although transmitted three times, it only requires one subframe to be received correctly for a new dialed digit or silence transition to be recognized.

A user receiving dialed digits should not filter the duration of transitions before reproducing them.
If despite triple redundancy one or more transitions are lost, a user shall continue to play out the preceding tone. It is the option to indicate the end of a tone if no further dialed digit packets are received within a period of 2 seconds.

Annex H
Channel Associated Signaling Bits

(This Annex forms an integral part of this Implementation Agreement)

H.1 General

This annex defines the packet format and procedures that shall be used to transport Channel Associated Signaling (CAS) bits over an MPLS connection in control subframe.

The concepts of CAS are defined in ITU-T Recommendation G.704 – see 3.1/G.704 for the 1544 kbit/s interface and 5.1/G.704 for the 2048 kbit/s interface.

The transmission of CAS control subframes is optional.

H.2 CAS control subframes

CAS control subframes are sent with triple redundant transmission (see § 7.4).

The format of CAS control subframes is shown in Figure H-1.

<table>
<tr><td colspan="10">0</td><td colspan="10">1</td><td colspan="10">2</td><td colspan="2">3</td><td></td></tr>
<tr><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>0</td><td>1</td><td></td></tr>
<tr><td colspan="8">CID value</td><td colspan="8">Payload type = 241</td><td colspan="16">Timestamp</td><td>0</td></tr>
<tr><td colspan="8">Redundancy</td><td colspan="4">Reserved</td><td>A</td><td>B</td><td>C</td><td>D</td><td colspan="16">Reserved</td><td>1</td></tr>
</table>

Figure H-1. CAS bits transfer frame format

The fields designated A, B, C and D contains the current value of the corresponding CAS bits.

H.3 Transmitter Procedures

When a transmitter wishes to convey to the receiver a change in state of the ABCD bits, the CAS control subframe shall be sent with triple redundancy at intervals of 5 ms.

Every 5 seconds thereafter the CAS control subframe is sent to refresh the ABCD state (with the Redundancy field coded as value 3).

Procedures to debounce CAS bits at the transmitter are outside the scope of this Implementation Agreement. Within certain limits, transient changes in the state of the CAS bits can be considered insignificant.

If a new state change is conveyed before the triple redundancy of a previous state change has completed, the transmitter shall stop sending control subframes for the previous state, in order to avoid the interleaving of two different time stamps.

If an external interface supplies fewer than four independent CAS bits, e.g. 1544 kbit/s with the 12-frame multiframe, the transmitting user shall aggregate and map the supplied bits to the four ABCD bits that the control subframe transfers. In particular, the sequence {A, B, A', B'} shall be transferred as C = A', D = B' and the sequence {A, A', A", A'"} as B = A', C = A", D = A'".

H.4 Receiver Procedures

If a receiver is interpreting the semantics of signaling, it should filter out (debounce) insignificant transient changes in the state of the CAS bits.

In order to make full use of triple redundancy without introducing extra delay variation, the receiver should wait before indicating changes in the state of the CAS bits until such time as it should have received all three copies of a transition. Although transmitted three times, it only requires one packet to be received correctly for a change in the state of the CAS bits to be recognized.