ETSI TS 126 114 V7.1.0 (2007-06)

Technical Specification

**Universal Mobile Telecommunications System (UMTS);
IP Multimedia Subsystem (IMS);
Multimedia telephony;
Media handling and interaction
(3GPP TS 26.114 version 7.1.0 Release 7)**

Reference
RTS/TSGS-0426114v710

Keywords
UMTS

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Intellectual Property Rights

IPRs essential or potentially essential to the present document may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (http://webapp.etsi.org/IPR/home.asp).

Pursuant to the ETSI IPR Policy, no investigation, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Foreword

This Technical Specification (TS) has been produced by ETSI 3rd Generation Partnership Project (3GPP).

The present document may refer to technical specifications or reports using their 3GPP identities, UMTS identities or GSM identities. These should be interpreted as being references to the corresponding ETSI deliverables.

The cross reference between GSM, UMTS, 3GPP and ETSI identities can be found under http://webapp.etsi.org/key/queryform.asp.

Contents

Foreword

This Technical Specification has been produced by the 3rd Generation Partnership Project (3GPP).

The contents of the present document are subject to continuing work within the TSG and may change following formal TSG approval. Should the TSG modify the contents of the present document, it will be re-released by the TSG with an identifying change of release date and an increase in version number as follows:

Version x.y.z

where:

x the first digit:

1 presented to TSG for information;

2 presented to TSG for approval;

3 or greater indicates TSG approved document under change control.

y the second digit is incremented for all changes of substance, i.e. technical enhancements, corrections, updates, etc.

z the third digit is incremented when editorial only changes have been incorporated in the document.

Introduction

Multimedia Telephony Service for IMS (MTSI), here also referred to as Multimedia Telephony, is a standardized IMS telephony service in 3GPP Release 7 that builds on the IMS capabilities already provided in 3GPP Releases 5 and 6. The objective of defining a service is to specify the minimum set of capabilities required in the IP Multimedia Subsystem to secure multi-vendor and multi-operator inter-operability for Multimedia Telephony and related Supplementary Services.

The user experience of multimedia telephony is expected to be equivalent to or better than corresponding circuit-switched telephony services. Multimedia telephony also exploits the richer capabilities of IMS. In particular, multiple media components can be used and dynamically added or dropped during a session.

1 Scope

The present document specifies a client for the Multimedia Telephony Service for IMS (MTSI) supporting conversational speech (including DTMF), video and text transported over RTP with the scope to deliver a user experience equivalent to or better than that of Circuit Switched (CS) conversational services using the same amount of network resources. It defines media handling (e.g. signalling, transport, jitter buffer management, packet-loss handling, adaptation), as well as interactivity (e.g. adding or dropping media during a call). The focus is to ensure a reliable and interoperable service with a predictable media quality, while allowing for flexibility in the service offerings.

The scope includes maintaining backward compatibility in order to ensure seamless inter-working with existing services available in the CS domain, such as CS speech and video telephony, as well as with terminals of earlier 3GPP releases. In addition, inter-working with traditional PSTN and emerging TISPAN network is covered.

The specification is written in a forward-compatible way in order to allow additions of media components and functionality in releases after Release 7.

NOTE 1: MTSI clients can support more than conversational speech, video and text, which is the scope of the present document. See 3GPP TS 22.173 [2] for the definition of the Multimedia Telephony Service for IMS.

NOTE 2: 3GPP TS 26.235 [3] and 3GPP TS 26.236 [4] do not include the specification of an MTSI client, although they include conversational multimedia applications. Only those parts of 3GPP TS 26.235 [3] and 3GPP TS 26.236 [4] that are specifically referenced by the present document apply to Multimedia Telephony Service for IMS.

NOTE 3: The present document was started as a conclusion from the study in 3GPP TR 26.914 [5] on optimization opportunities in Multimedia Telephony for IMS (3GPP TR 22.973 [6]).

2 References

The following documents contain provisions which, through reference in this text, constitute provisions of the present document.

- References are either specific (identified by date of publication, edition number, version number, etc.) or non-specific.
- For a specific reference, subsequent revisions do not apply.
- For a non-specific reference, the latest version applies. In the case of a reference to a 3GPP document (including a GSM document), a non-specific reference implicitly refers to the latest version of that document *in the same Release as the present document*.

[1] 3GPP TR 21.905: "Vocabulary for 3GPP Specifications".

[2] 3GPP TS 22.173: "IP Multimedia Core Network Subsystem (IMS) Multimedia Telephony Service and supplementary services; Stage 1".

[3] 3GPP TS 26.235: "Packet switched conversational multimedia applications; Default codecs".

[4] 3GPP TS 26.236: "Packet switched conversational multimedia applications; Transport protocols".

[5] 3GPP TR 26.914: "Multimedia telephony over IP Multimedia Subsystem (IMS); Optimization opportunities".

[6] 3GPP TR 22.973: "IMS Multimedia Telephony service; and supplementary services".

[7] 3GPP TS 24.229: "IP multimedia call control protocol based on Session Initiation Protocol (SIP) and Session Description Protocol (SDP); Stage 3".

[8] IETF RFC 4566 (2006): "SDP: Session Description Protocol", M. Handley, V. Jacobson and C. Perkins.

[9] IETF RFC 3550 (2003): "RTP: A Transport Protocol for Real-Time Applications", H. Schulzrinne, S. Casner, R. Frederick and V. Jacobson.

[10] IETF RFC 3551 (2003): "RTP Profile for Audio and Video Conferences with Minimal Control", H. Schulzrinne and S. Casner.

[11] 3GPP TS 26.071: "Mandatory Speech Codec speech processing functions; AMR Speech CODEC; General description".

[12] 3GPP TS 26.090: "Mandatory Speech Codec speech processing functions; Adaptive Multi-Rate (AMR) speech codec; Transcoding functions".

[13] 3GPP TS 26.073: "ANSI C code for the Adaptive Multi Rate (AMR) speech codec".

[14] 3GPP TS 26.104: "ANSI-C code for the floating-point Adaptive Multi Rate (AMR) speech codec".

[15] 3GPP TS 26.093: "Mandatory speech codec speech processing functions; Adaptive Multi-Rate (AMR) speech codec; Source controlled rate operation".

[16] 3GPP TS 26.103: "Speech codec list for GSM and UMTS".

[17] 3GPP TS 26.171: "Speech codec speech processing functions; Adaptive Multi-Rate - Wideband (AMR-WB) speech codec; General description".

[18] 3GPP TS 26.190: "Speech codec speech processing functions; Adaptive Multi-Rate - Wideband (AMR-WB) speech codec; Transcoding functions".

[19] 3GPP TS 26.173: "ANCI-C code for the Adaptive Multi Rate - Wideband (AMR-WB) speech codec".

[20] 3GPP TS 26.204: "Speech codec speech processing functions; Adaptive Multi-Rate - Wideband (AMR-WB) speech codec; ANSI-C code".

[21] 3GPP TS 26.193: "Speech codec speech processing functions; Adaptive Multi-Rate - Wideband (AMR-WB) speech codec; Source controlled rate operation".

[22] ITU-T Recommendation H.263 (2005): "Video coding for low bit rate communication".

[23] ISO/IEC 14496-2:2004: "Information technology - Coding of audio-visual objects - Part 2: Visual".

[24] ITU-T Recommendation H.264 (2005): "Advanced video coding for generic audiovisual services" | ISO/IEC 14496-10:2005: "Information technology - Coding of audio-visual objects - Part 10: Advanced Video Coding".

[25] IETF RFC 3984 (2005): "RTP Payload Format for H.264 Video", S. Wenger, M.M. Hannuksela, T. Stockhammer, M. Westerlund and D. Singer.

[26] ITU-T Recommendation T.140 (1998): "Protocol for multimedia application text conversation".

[27] ITU-T Recommendation T.140 (2000): "Protocol for multimedia application text conversation - Addendum 1".

[28] IETF RFC 4867 (2007): "RTP Payload Format and File Storage Format for the Adaptive Multi-Rate (AMR) and Adaptive Multi-Rate Wideband (AMR-WB) Audio Codecs", J. Sjoberg, M. Westerlund, A. Lakaniemi and Q. Xie.

[29] IETF RFC 4629 (2007): "RTP Payload Format for ITU-T Rec. H.263 Video", J. Ott, C. Bormann, G. Sullivan, S. Wenger and R. Even.

[30] IETF RFC 3016 (2000): "RTP Payload Format for MPEG-4 Audio/Visual Streams", Y. Kikuchi, T. Nomura, S. Fukunaga, Y. Matsui and H. Kimata.

[31] IETF RFC 4103 (2005): "RTP Payload for Text Conversation", G. Hellstrom and P. Jones.

[32] IETF RFC 3555 (2003): "MIME Type Registration of RTP Payload Formats", S. Casner and P. Hoschka.

[33] 3GPP TR 25.993: "Typical examples of Radio Access Bearers (RABs) and Radio Bearers (RBs) supported by Universal Terrestrial Radio Access (UTRA)".

[34] 3GPP TS 22.105: "Services and service capabilities".

[35] 3GPP TS 26.131: "Terminal acoustic characteristics for telephony; Requirements".

[36] 3GPP TS 26.132: "Speech and video telephony terminal acoustic test specification".

[37] 3GPP TS 28.062: "Inband Tandem Free Operation (TFO) of speech codecs; Service description; Stage 3".

[38] 3GPP TS 23.153: "Out of band transcoder control; Stage 2".

[39] IETF RFC 0768 (1980): "User Datagram Protocol", J. Postel.

[40] IETF RFC 4585 (2006): "Extended RTP Profile for Real-time Transport Control Protocol (RTCP) - Based Feedback (RTP/AVPF)", J. Ott, S. Wenger, N. Sato, C. Burmeister and J. Rey.

[41] RTP Tools: http://www.cs.columbia.edu/IRT/software/rtptools/.

[42] IETF RFC 3556 (2003): "Session Description Protocol (SDP) Bandwidth Modifiers for RTP Control Protocol (RTCP) Bandwidth", S. Casner.

[43] IETF Internet draft (2006): "Codec Control Messages in the RTP Audio-Visual Profile with Feedback (AVPF)", S. Wenger, U. Chandra, M. Westerlund and B. Burman, http://www.ietf.org/internet-drafts/draft-ietf-avt-avpf-ccm-03.txt.

[44] 3GPP TR 29.863: "Feasibility Study for the multimedia inter-working between the IM CN subsystem and CS Networks".

[45] 3GPP TS 26.111: "Codec for circuit switched multimedia telephony service; Modifications to H.324".

[46] 3GPP TS 23.172: "Technical realization of Circuit Switched (CS) multimedia service; UDI/RDI fallback and service modification; Stage 2".

[47] 3GPP TS 23.002: "Network Architecture".

[48] IETF RFC 3388 (2002): "Grouping of Media Lines in the Session Description Protocol (SDP)", G. Camarillo, G. Eriksson, J. Holler and H. Schulzrinne.

[49] IETF RFC 4102 (2005): "Registration of the text/red MIME Sub-Type", P. Jones.

[50] ITU-T H.248: "Packages for text conversation, fax and call discrimination".

[51] ETSI EG 202 320: "Duplex Universal Text and Voice".

[52] 3GPP TS 26.226: "Cellular text telephone modem; General description".

[53] IETF RFC 4504 (2006): "SIP Telephony Device Requirements and Configuration", H. Sinnreich, Ed., S. Lass and C. Stredicke.

[54] ITU-T Recommendation V.151: "Procedures for end-to-end connection of analogue PSTN text telephones over an IP network utilizing text relay".

[55] ITU-T Recommendation V.152: "Procedures for supporting Voice Band Data over IP networks".

[56] IETF RFC 3448 (2003): "TCP Friendly Rate Control (TFRC): Protocol Specification", M. Handley, S. Floyd, J. Padhye and J. Widmer.

[57] 3GPP TS 24.173: "IMS Multimedia Telephony Communication Service and Supplementary Services".

[58] IETF RFC 3264 (2002): "An Offer/Answer Model with the Session Description Protocol (SDP)", J. Rosenberg and H. Schulzrinne.

[59] 3GPP TS 26.141: "IP Multimedia System (IMS) Messaging and Presence; Media formats and codecs".

[60] 3GPP TS 26.234: "Transparent end-to-end Packet-switched Streaming Service; Protocols and codecs".

[61] IETF RFC 4733 (2006): "RTP Payload for DTMF Digits, Telephony Tones, and Telephony Signals", H. Schulzrinne and T.Taylor.

[62] 3GPP TS 23.014. "Support of Dual Tone Multi-Frequency (DTMF) signalling",

[63] ETSI ES 201 235-2, v1.2.1: "Specification of Dual Tone Multi-Frequency (DTMF); Transmitters and Receivers; Part 2: Transmitters".

[64] 3GPP TS 23.107: "Quality of Service (QoS) concept and architecture".

3 Definitions and abbreviations

3.1 Definitions

For the purposes of the present document, the terms and definitions given in 3GPP TR 21.905 [1] and the following apply:

NOTE: A term defined in the present document takes precedence over the definition of the same term, if any, in 3GPP TR 21.905 [1].

example: text used to clarify abstract rules by applying them literally

3.2 Abbreviations

For the purposes of the present document, the abbreviations given in 3GPP TR 21.905 [1] and the following apply:

NOTE: An abbreviation defined in the present document takes precedence over the definition of the same abbreviation, if any, in 3GPP TR 21.905 [1].

AL-SDU Application Layer - Service Data Unit
AMR Adaptive Multi-Rate
AMR-NB Adaptive Multi-Rate - NarrowBand
AMR-WB Adaptive Multi-Rate - WideBand
APP APPlication-defined RTCP packet
ARQ Automatic repeat ReQuest
AS Application Server
AVC Advanced Video Coding
CCM Codec Control Messages
CDF Cumulative Distribution Function
CMR Codec Mode Request
cps characters per second
CS Circuit Switched
CSCF Call Session Control Function
CTM Cellular Text telephone Modem
DTMF Dual Tone Multi-Frequency
DTX Discontinuous Transmission
GIP Generic IP access
GOB Group Of Blocks
H-ARQ Hybrid - ARQ
HSPA High Speed Packet Access
IDR Instantaneous Decoding Refresh
IMS IP Multimedia Subsystem
IP Internet Protocol
IPv4 Internet Protocol version 4

ITU-T	International Telecommunications Union - Telecommunications
JBM	Jitter Buffer Management
MGCF	Media Gateway Control Function
MGW	Media GateWay
MIME	Multipurpose Internet Mail Extensions
MPEG	Moving Picture Experts Group
MRFP	Media Resource Function Processor
MTSI	Multimedia Telephony Service for IMS
MTU	Maximum Transfer Unit
NACK	Negative ACKnowledgment
NTP	Network Time Protocol
PDP	Packet Data Protocol
PLI	Picture Loss Indication
POI	Point Of Interconnect
PSTN	Public Switched Telephone Network
QoS	Quality of Service
RoHC	Robust HeaderCompression
RR	Receiver Report
RTCP	RTP Control Protocol
RTP	Real-time Transport Protocol
SDP	Session Description Protocol
SID	SIlence Descriptor
SIP	Session Initiation Protocol
SR	Sender Report
TFO	Tandem-Free Operation
TISPAN	Telecoms and Internet converged Services and Protocols for Advanced Network
TMMBN	Temporary Maximum Media Bit-rate Notification
TMMBR	Temporary Maximum Media Bit-rate Request
TrFO	Transcoder-Free Operation
UDP	User Datagram Protocol
UE	User Equipment
VoIP	Voice over IP

4 System description

4.1 System

A Multimedia Telephony Service for IMS call uses the Call Session Control Function (CSCF) mechanisms to route control-plane signalling between the UEs involved in the call (see figure 4.1). In the control plane, Application Servers (AS) should be present and may provide supplementary services such as call hold/resume, call forwarding and multi-party calls, etc.

The scope of the present document is to specify the media path. In the example in figure 4.1, it is routed directly between the GGSNs outside the IMS.

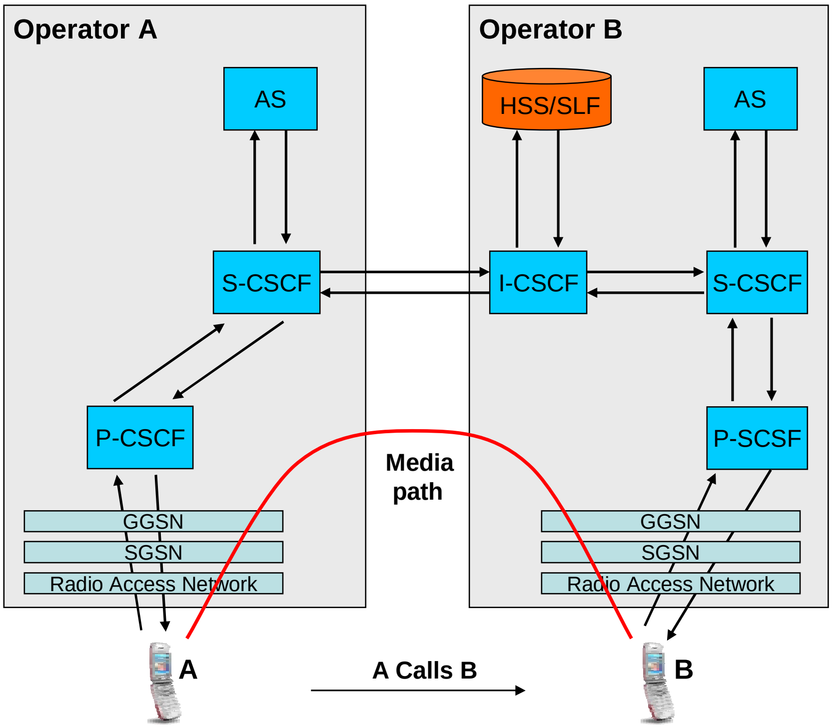
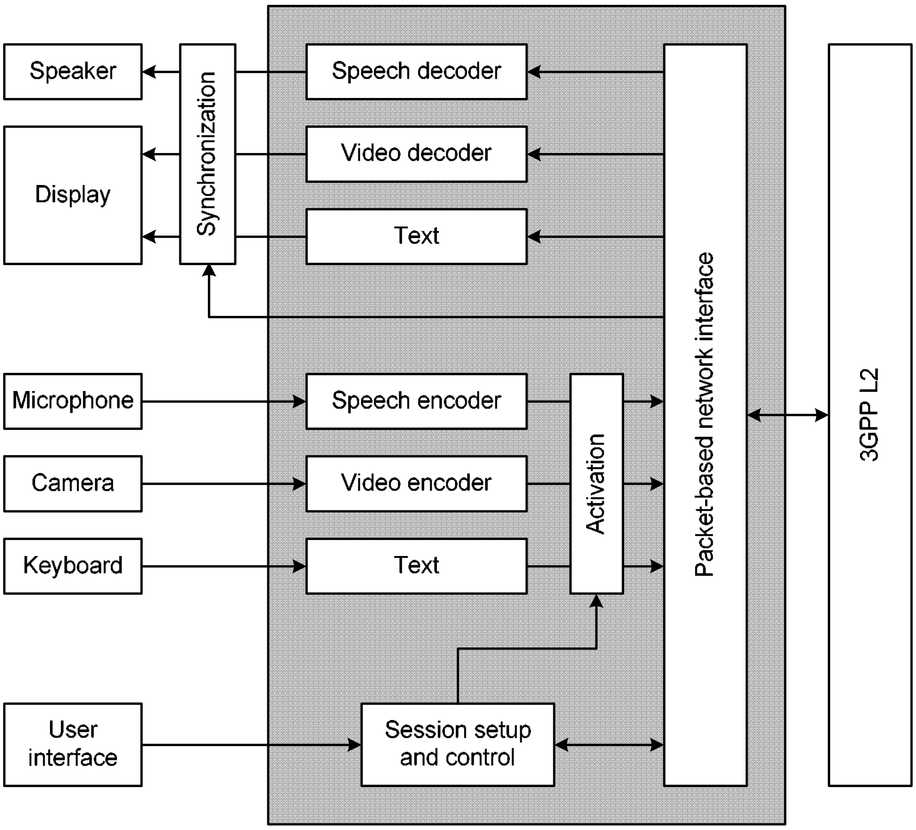


Figure 4.1: High-level architecture figure showing the nodes involved in an MTSI call set-up

4.2 Client

The functional components of an MTSI client are shown in figure 4.2.



NOTE: The grey box marks the scope of the present document.

Figure 4.2: Functional components of an MTSI client

The scope of the present document is to specify media handling and interaction, which includes media control, media codecs, as well as transport of media and control data. General control-related elements of an MTSI client, such as SIP signalling (3GPP TS 24.229 [7]), fall outside this scope, albeit parts of the session setup handling and session control are defined here:

- usage of SDP (RFC 4566 [8]) in SIP invitations for capability negotiation and media stream setup.
- set-up and control of the individual media streams between clients. It also includes interactivity, such as adding and dropping of media components.

Transport of media consists of the encapsulation of the coded media in a transport protocol as well as handling of coded media received from the network. This is shown in figure 4.2 as the "packet based network interface" and is displayed in more detail in the user-plane protocol stack in figure 4.3. The basic MTSI client defined here specifies media codecs for speech, video and text (see clause 5). All media components are transported over RTP with each respective payload format mapped onto the RTP (RFC 3550 [9]) streams.

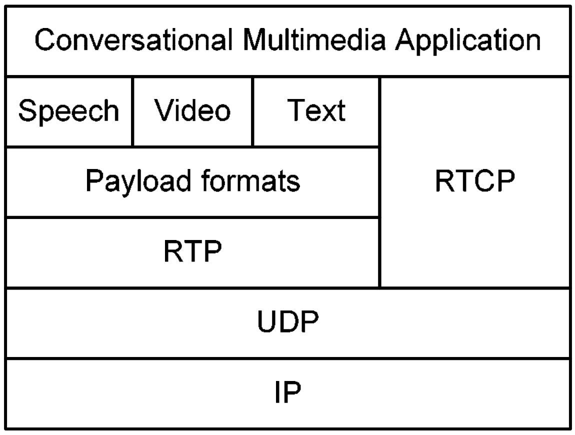


Figure 4.3: User plane protocol stack for a basic MTSI capable terminal

4.3 MRFP and MGW

A Media Resource Function Processor (MRFP), see 3GPP TS.23.002 [47], may be inserted in the media path for certain supplementary services (e.g. conference) and to provide transcoding.

A Media Gateway (MGW), see 3GPP TS 23.002 [47], may be used to provide inter-working between different networks and services. For example, a MGW may provide inter-working between MTSI and 3G-324M services. The MGW may have more limited functionality than a MTSI client, e.g. when it comes to the supported bitrates of media. The inter-working aspects are described in more detail in clause 12.

5 Media codecs

5.1 Media components

The Multimedia Telephony Service for IMS supports simultaneous transfer of multiple media components with real-time characteristics. Media components denote the actual components that the end-user experiences.

The following media components are considered as core components. At least one of these components is present in all conversational multimedia telephony sessions.

- **Speech:** The sound that is picked up by a microphone and transferred from terminal A to terminal B and played out in an earphone/loudspeaker. Speech includes detection and generation of DTMF signals.
- **Video:** The moving image that is captured by a camera of terminal A and rendered on the display of terminal B.

- **Text:** The characters typed on a keyboard or drawn on a screen on terminal A and rendered in real time on the display of terminal B. The flow is time-sampled so that no specific action is needed from the user to request transmission.

The above core media components are transported in real time from one terminal to the other using RTP (RFC 3550 [9]). All media components can be added or dropped during an ongoing session as required either by the end-user or by controlling nodes in the network, assuming that when adding components, the capabilities of the UE support the additional component.

NOTE: The terms voice and speech are synonyms. The present document uses the term speech.

5.2 Codecs for terminals

5.2.1 Speech

MTSI terminals offering speech communication shall support:

- AMR speech codec (3GPP TS 26.071 [11], 3GPP TS 26.090 [12], 3GPP TS 26.073 [13] and 3GPP TS 26.104 [14]) including all 8 modes and source controlled rate operation 3GPP TS 26.093 [15]. The terminal shall be capable of operating with any subset of these 8 codec modes.

The codec mode set Config-NB-Code=1 (3GPP TS 26.103 [16]) {AMR-NB12.2, AMR-NB7.4, AMR-NB5.9 and AMR-NB4.75} should be used unless the session-setup negotiation determines that other codec modes shall be used.

When transmitting, the terminal shall be capable of aligning codec mode changes to every frame border, and shall also be capable of restricting codec mode changes to be aligned to every other frame border, e.g. like UMTS_AMR_2 (3GPP TS 26.103 [16]). The terminal shall also be capable of restricting codec mode changes to neighbouring codec modes within the negotiated codec mode set. When receiving, the terminal shall allow codec mode changes at any frame border and to any codec mode within the negotiated codec mode set.

MTSI terminals offering wideband speech communication at 16 kHz sampling frequency shall support:

- AMR wideband codec (3GPP TS 26.171 [17], 3GPP TS 26.190 [18], 3GPP TS 26.173 [19] and 3GPP TS 26.204 [20]) including all 9 modes and source controlled rate operation 3GPP TS 26.193 [21]. The terminal shall be capable of operating with any subset of these 9 codec modes.

The codec mode set Config-WB-Code=0 (3GPP TS 26.103 [16]) {AMR-WB12.65, AMR-WB8.85 and AMR-WB6.60} should be used unless the session-setup negotiation determines that other codec modes shall be used.

When transmitting, the terminal shall be capable of aligning codec mode changes to every frame border, and shall also be capable of restricting codec mode changes to be aligned to every other frame border, e.g. like UMTS_AMR_WB (3GPP TS 26.103 [16]). The terminal shall also be capable of restricting codec mode changes to neighbouring codec modes within the negotiated codec mode set. When receiving, the terminal shall allow codec mode changes at any frame border and to any codec mode within the negotiated codec mode set.

MTSI terminals offering wideband speech communication shall also offer narrowband speech communications. When offering both wideband speech and narrowband speech communication, wideband shall be listed as the first payload type in the m line of the SDP offer (RFC 4566 [8]).

Encoding of DTMF is described in Annex G.

5.2.2 Video

MTSI terminals offering video communication shall support:

- ITU-T Recommendation H.263 [22] Profile 0 Level 45.

In addition they should support:

- ITU-T Recommendation H.263 [22] Profile 3 Level 45;

- MPEG-4 (Part 2) Visual [23] Simple Profile Level 3with the following constraints:

- Number of Visual Objects supported shall be limited to 1.
- The maximum frame rate shall be 30 frames per second.
- The maximum f_code shall be 2.
- The intra_dc_vlc_threshold shall be 0.
- The maximum horizontal luminance pixel resolution shall be 352 pels/line.
- The maximum vertical luminance pixel resolution shall be 288 pels/VOP.
- If AC prediction is used, the following restriction applies: QP value shall not be changed within a VOP (or within a video packet if video packets are used in a VOP). If AC prediction is not used, there are no restrictions to changing QP value.

- ITU-T Recommendation H.264 / MPEG-4 (Part 10) AVC [24] Baseline Profile Level 1.1 without requirements on output timing conformance (annex C of [24]). Each sequence parameter set of H.264 (AVC) shall contain the vui_parameters syntax structure including the num_reorder_frames syntax element set equal to 0.

The H.264 (AVC) decoder in a multimedia terminal shall either start decoding immediately when it receives data (even if the stream does not start with an IDR access unit) or alternatively no later than it receives the next IDR access unit or the next recovery point SEI message, whichever is earlier in decoding order. The decoding process for a stream not starting with an IDR access unit shall be the same as for a valid H.264 (AVC) bit stream. However, the client shall be aware that such a stream may contain references to pictures not available in the decoded picture buffer. The display behaviour of the client is out of scope of the present document.

NOTE 1: If a codec is supported at a certain level, then all (hierarchically) lower levels shall be supported as well. Examples of lower levels include Level 10 for H.263 Profile 0 and 3, Level 0 for MPEG-4 Visual Simple Profile and Level 1 for H.264 (AVC) Baseline Profile. However, as for instance Level 20 is not hierarchically lower than Level 45 of H.263 Profile 0 and 3, support for Level 45 does not imply support for Level 20.

NOTE 2: All levels are minimum requirements. Higher levels may be supported and used for negotiation.

NOTE 3: Terminals may use full-frame freeze and full-frame freeze release SEI messages of H.264 (AVC) to control the display process.

NOTE 4: An H.264 (AVC) encoder should code redundant slices only if it knows that the far-end decoder makes use of this feature (which is signalled with the redundant-pic-cap MIME/SDP parameter as specified in RFC 3984 [25]). H.264 (AVC) encoders should also pay attention to the potential implications on end-to-end delay.

NOTE 5: If a codec is supported at a certain level, it implies that on the receiving side, the decoder is required to support the decoding of bitstreams up to the maximum capability of this level. On the sending side, the support of a particular level does not imply that the encoder may produce a bitstream up to the maximum capability of the level.

5.2.3 Real-time text

MTSI terminals offering real time text conversation shall support:

- ITU-T Recommendation T.140 [26] and [27].

T.140 specifies coding and presentation features of real-time text usage. Text characters are coded according to the UTF-8 transform of ISO 10646-1 (Unicode).

A minimal subset of the Unicode character set, corresponding to the Latin-1 part shall be supported, while the languages in the regions where the terminal is intended to be used should be supported.

Presentation control functions from ISO 6429 are allowed in the T.140 media stream. A mechanism for extending control functions is included in ITU-T Recommendation T.140 [26] and [27]. Any received non-implemented control code must not influence presentation.

A terminal shall store the conversation in a presentation buffer during a call for possible scrolling, saving, display re-arranging, erasure, etc. At least 800 characters shall be kept in the presentation buffer during a call.

Note that erasure (backspace) of characters is included in the T.140 editing control functions. It shall be possible to erase all characters in the presentation buffer. The display of the characters in the buffer shall also be impacted by the erasure.

6 Media configuration

6.1 General

MTSI uses SIP and SDP for media negotiation and configuration. General SIP signalling and session setup for IMS are defined in 3GPP TS 24.229 [7], whereas this clause specifies SDP usage and media handling specifically for MTSI, including offer/answer considerations in the capability negotiation.

6.2 Session setup procedures

6.2.1 General

The session setup shall determine for each media: UDP port number(s); codec(s); RTP Payload Type number(s), RTP Payload Format(s) and any additional session parameters.

An MTSI terminal shall only offer a single RTP profile per media stream. This profile shall be the most suitable for the media, see below for further recommendations for each media type. The MTSI terminal shall accept both AVP and AVPF offers in order to support interworking. If an MTSI terminal gets a media or the complete session rejected when using AVPF, it should re-invite replacing all AVPF with AVP on all media lines where it did not receive explicit indication that AVPF was accepted.

6.2.2 Speech

For AMR or AMR-WB encoded media, the session setup shall determine: if all codec modes can be used or if the operation needs to be restricted to a subset; if the bandwidth-efficient payload format can be used or if the octet-aligned payload format must be used; if codec mode changes shall be restricted to be aligned to only every other frame border or if codec mode changes can occur at any frame border; if codec mode changes must be restricted to only neighbouring modes within the negotiated codec mode set or if codec mode changes can be performed to any mode within the codec mode set; the number of speech frames that should be encapsulated in each RTP packet and the maximum number of speech frames that may be encapsulated in each RTP packet.

If the session setup negotiation concludes that multiple configuration variants are possible in the session then the default operation should be used as far as the agreed parameters allow, see clause 7.5.2.1. It should be noted that the default configurations are slightly different for different access types.

An MTSI terminal offering a speech media session for narrow-band speech and/or wide-band speech should offer SDP according to the examples in clauses A.1 to A.3.

An MTSI terminal shall offer AVPF for speech media streams. An MTSI terminal may offer AVP if RTCP is not used.

Session setup for sessions including speech and DTMF events is described in Annex G.

6.2.3 Video

If video is used in a session, the session setup shall determine video codec, profile and level.

An MTSI terminal shall offer AVPF for all media streams containing video.

Examples of SDP offers and answers for video can be found in clause A.4.

NOTE: For H.264 / MPEG-4 (Part 10) AVC, the optional max-rcmd-nalu-size receiver-capability parameter of RFC 3984 [25] should be set to the smaller of the MTU size (if known) minus header size or 1 400 bytes (otherwise).

6.2.4 Text

An MTSI terminal should offer AVP for all media streams containing text. Only in cases where there is an explicit demand for the AVPF RTCP reporting timing or feedback messages AVPF shall be used.

Examples of SDP offers for text can be found in clause A.5.

6.2.5 Bandwidth negotiation

The SDP shall include bandwidth information for each media stream and also for the session in total. The bandwidth information for each media stream and for the session is defined by the Application Specific (AS) bandwidth modifier as defined in RFC 4566 [8].

SDP examples incorporating bandwidth modifiers are shown in annex A.

6.2.6 The Synchronization Info attribute "3gpp_sync_info"

Synchronization jitter (also known as synchronization or inter-media skew) is defined as the amount of synchronization delay between media streams that needs to be maintained during the synchronization process (at the receiver side), which is acceptable to a session (or the sender of the multimedia streams) for a good user experience.

Tight synchronization between the constituent streams is not necessary for all types of MTSI sessions. For instance, during a VoIP call, one of the call participants may wish to share a video clip or share his/her camera view. In this situation, the sender may want to relax the requirement on the receiver to synchronize the audio and the video streams in order to maintain a good video quality without stressing on tight audio/video synchronization. The Synchronization Info attribute defined in the present document is not just limited to lip-sync between audio/video streams, but is also applicable to any two media streams that need to be synchronized during an MTSI session. This attribute allows an MTSI terminal to specify whether or not media streams should be synchronized. In case the choice is to have synchronization between different streams, it is up to the implementation, use case and application to decide the exact amount of synchronization jitter allowed between the streams to synchronize.

The ABNF for the synchronization info attribute is described as follows:

```
Synchronization-Info    = "a" "=" "3gpp_sync_info" ":" sync-value

sync-value              = "Sync" / "No Sync"
```

The value "Sync" indicates that synchronization between media shall be maintained. The value "No Sync" indicates that No Synchronization is required between the media.

The parameter "3gpp_sync_info" should be included in the SDP at the session level and/or at the media level. Its usage is governed by the following rules:

1. At the session level, the "3gpp_sync_info" attribute shall be used with the group attribute defined in RFC 3388 [48]. The group attribute indicates to the receiver which streams (identified by their mid attributes) that are to be synchronized. The "3gpp_sync_info" attribute shall follow the "group: LS" line in the SDP.

2. At the media level, the "3gpp_sync_info" attribute shall assume a value of "No Sync" only. It indicates to the receiver that this particular media stream is not required to be synchronized with any other media stream in the session. The use of the "mid" attribute of RFC 3388 [48] is optional in this case. If the "mid" attribute is used for any other media in the session, then "mid" with this media line shall be used also according to RFC 3388 [48]. Otherwise, it is not necessary to tie the "3gpp_sync_info" attribute with the "mid" attribute.

3. When the "3gpp_sync_info" attribute is defined at both session level (with the "group" attribute) and media level, then the media level attribute shall override the session level attribute. Thus if the "3gpp_sync_info" attribute is defined at the media level, then that particular media stream is not to be synchronized with any other media stream in the session (even if the "3gpp_sync_info" is defined at the session level for this media stream).

The calling party (or the initiator or offerer of the multimedia stream) should include the "3gpp_sync_info" attribute in the SDP which is carried in the initial INVITE message. Upon reception of the INVITE message that includes the "3gpp_sync_info" attribute, the other party in the session should include its own "3gpp_sync_info" attribute (with its own wish for synchronization or no synchronization) in the 200/OK response message.

There are no offer/answer implications on the "3gpp_sync_info" attribute. The "3gpp_sync_info" attribute in the calling party SDP is only an indication to the called party of the synchronization requirement between streams that should be maintained. Similarly the "3gpp_sync_info" attribute value from the called party is an indication to the calling party of the synchronization requirements between specified media streams. The "3gpp_sync_info" attribute value can be different for the calling and the called parties.

SDP examples using the "3gpp_sync_info" attribute are given in clause A.7.

NOTE: Default operation in the absence of the "3gpp_sync_info" attribute in SDP is to maintain synchronization between media streams.

6.2.7 Negotiated QoS parameters

The term "negotiated" in the present document describes the end result of a QoS negotiation between an MTSI terminal and the network (or the end result of what the network grants to the terminal even if no negotiation takes place).

In case an MTSI terminal is made aware that the value of the negotiated Guaranteed Bit Rate differs from the b=AS bandwidth modifier attribute during the initial session setup in an MTSI terminal (sender or receiver), the MTSI terminal shall send to the other party the negotiated Guaranteed Bit Rate via the SIP UPDATE method using the b=AS bandwidth modifier attribute. The other MTSI party (receiver or sender) shall respond by sending its known negotiated Guaranteed Bit Rate via the SIP 200/OK response to the UPDATE message.

Any subsequent QoS changes indicated to the MTSI terminal during an MTSI session (including the cases described in Clause 10.3) shall be signalled by the MTSI party (subject to the QoS update procedure) to the other MTSI party using the same signalling described above.

Examples of SDP using negotiated QoS are given in clause A.8.

6.3 Session control procedures

During session renegotiation for adding or removing media components, the SDP offerer should continue to use the same media (m=) line(s) from the previously negotiated SDP for the media components that are not being added or removed.

7 Data transport

7.1 General

MTSI terminals shall support an IP-based network interface for the transport of session control and media data. Control-plane signalling is sent using SIP; see 3GPP TS 24.229 [7] for further details. User plane media data is sent over RTP/UDP/IP. An overview of the user plane protocol stack can be found in figure 4.3 of the present document.

7.2 RTP profiles

MTSI terminals shall transport speech, video and real-time text using RTP (RFC 3550 [9]) over UDP (RFC 0768 [39]). The following profiles of RTP shall be supported:

- RTP Profile for Audio and Video Conferences with Minimal Control (RFC 3551 [10]), also called RTP/AVP;
- Extended RTP Profile for RTCP-based Feedback (RTP/AVPF) (RFC 4585 [40]), also called RTP/AVPF.

The support of AVPF requires an MTSI terminal to implement the RTCP transmission rules, the signalling mechanism for SDP and the feedback messages explicitly mentioned in the present document.

7.3 RTCP usage

7.3.1 General

The RTP implementation shall include an RTCP implementation.

The bandwidth for RTCP traffic shall be described using the "RS" and "RR" SDP bandwidth modifiers at media level, as specified by RFC 3556 [42]. Therefore, an MTSI terminal shall include the "b=RS:" and "b=RR:" fields in SDP, and shall be able to interpret them. There shall be a limit on the allowed RTCP bandwidth for an RTP session signalled by the terminal. This limit is defined as follows:

- 4 000 bps for the RS field (at media level);
- 3 000 bps for the RR field (at media level).

If the session described in the SDP is a point-to-point speech only session, the UE may request the deactivation of RTCP by setting its RTCP bandwidth modifiers to zero.

If a UE receives SDP bandwidth modifiers for RTCP equal to zero from the originating UE, it should reply (via the SIP protocol) by setting its RTCP bandwidth using SDP bandwidth modifiers with values equal to zero.

RTCP packets should be sent for all types of multimedia sessions to enable synchronization with other RTP transported media, remote end-point aliveness information, monitoring of the transmission quality, and carriage of feedback messages such as TMMBR for video and RTCP APP for speech. Point-to-point speech only sessions may not require these functionalities and may therefore turn off RTCP by setting the SDP bandwidth modifiers (RR and RS) to zero. When RTCP is turned off (for point-to-point speech only sessions) and the media is put on hold, the terminal should re-negotiate the RTCP bandwidth with SDP bandwidth modifiers values greater than zero, and send RTCP packets to the other end. This allows the remote end to detect link aliveness during hold. When media is resumed, the resuming terminal should turn off the RTCP sending again through a re-negotiation of the RTCP bandwidth with SDP bandwidth modifiers equal to zero.

When RTCP is turned off (for point-to-point speech only sessions) and if sending of an additional associated RTP stream becomes required and both RTP streams need to be synchronized, or if transport feedback due to lack of end-to-end QoS guarantees is needed, a terminal should re-negotiate the bandwidth for RTCP by sending an SDP with the RS bandwidth modifier greater than zero.

NOTE 1: Deactivating RTCP will disable the adaptation mechanism for speech defined in clause 10.2.

7.3.2 Speech

MTSI terminals offering speech shall support AVPF (RFC 4585 [40]) configured to operate in early mode. When allocating RTCP bandwidth, it is recommended to set the "b=RR:" and the "b=RS:" parameters to 5 % of the total session bandwidth. The value of "trr-int" should be set to zero or not transmitted at all (in which case the default "trr-int" value of zero will be assumed) when non-compound RTCP (see clause 7.3.5) is not used.

For speech sessions it is beneficial to keep the size of RTCP packets as small as possible in order to reduce the potential disruption of RTCP onto the RTP stream in bandwidth-limited channels. RTCP packet sizes can be minimized by using non-compound packets or using the parts of RTCP compound packets (according to RFC 3550 [9]) which are required by the application. RTCP compound packet sizes should be at most as large as 1 time and, at the same time, shall be at most as large as 4 times the size of the RTP packets (including UDP/IP headers) corresponding to the highest bit rate of the speech codec modes used in the session. RTCP non-compound and semi-compound packet sizes should be at most as large as 1 time and, at the same time, shall be at most as large as 2 times the size of the RTP packets (including UDP/IP headers) corresponding to the highest bit rate of the speech codec modes used in the session.

For speech, RTCP APP packets are used for adaptation (see clause 10.2).

7.3.3 Video

MTSI terminals offering video shall support AVPF (RFC 4585 [40]) configured to operate in early mode. The behaviour can be controlled by allocating enough RTCP bandwidth using "b=RR:" and "b=RS:" (see section 7.3.1) and setting the value of "trr-int".

MTSI terminals offering video shall support transmission and reception of AVPF NACK messages, as an indication of non-received media packets. MTSI terminals offering video shall also support reception of AVPF Picture Loss Indication (PLI). An MTSI terminal receiving NACK or PLI should take appropriate action to improve the situation for the terminal that sent NACK or PLI, although no action is mandated nor specified.

The Temporary Maximum Media Bit-rate Request (TMMBR) and Temporary Maximum Media Bit-rate Notification (TMMBN) messages of Codec-Control Messages (CCM) [43] shall be supported by MTSI terminals supporting video. See clause 10.3 for usage and clause B.1 for an example of bitrate adaptation.

7.3.4 Real-time text

For real-time text, RTCP reporting should be used according to general recommendations for RTCP.

7.3.5 Non-compound RTCP

MTSI terminals should support the use of non-compound RTCP reports [61]. A non-compound RTCP packet is an RTCP packet that does not follow the sending rules outlined in RFC 3550 [9] in the aspect that it does not necessarily contain the mandated RR/SR report blocks and SDES CNAME items.

If non-compound RTCP packets are supported, the following requirements apply on the RTCP receiver:

- The RTCP receiver shall be capable of parsing and decoding report blocks of the RTCP packet correctly even though some of the items mandated by RFC3550 [9] are missing.
- An SDP attribute ”ncp” is used to enable non-compound RTCP. This attribute shall be offered in SDP when the offer includes an offer for using the AVPF profile, see Annex A.9. A receiver that accepts the use of non-compound RTCP shall include the attribute in the SDP answer. If this attribute is not set in offer/answer, non-compound RTCP shall not be used in any direction.

If non-compound RTCP packets are supported , an RTCP sender transmitting non-compound RTCP packets shall follow the requirements listed below:

- AVPF early or immediate mode shall be used according to RFC4585 [40].
- Non-compound RTCP packets should be used for speech sessions, for transmission of adaptation feedback messages as defined in section 10.2 of this specification, or for transmission of regular feedback as individual non-compound RTCP packets (SR/RR, SDES or other APP packets). When regular feedback packets are transmitted, the individual packets that would belong to a compound RTCP packet shall be transmitted in a serial fashion, although adaptation feedback packets shall take precedence.
- Two or more non-compound RTCP individual packets should be stacked together, within the limits allowed by the maximum size of non-compound packets (see clause 7.3.2) (i.e., to form a semi-compound RTCP packet which is smaller than a compound RTCP packet).
- Compound RTCP packets with an SR/RR report block and CNAME SDES item should be transmitted on a regular basis as outlined in RFC 3550 [9] and RFC 4585 [40]. In order to control the allocation of bandwidth between non-compound RTCP and compound RTCP, the AVPF “trr-int” parameter should be used to set the minimum report interval for compound RTCP packets.
- The first transmitted RTCP packet shall be a compound RTCP packet as defined in RFC3550 [9] without the size restrictions defined in clause 7.3.2.

The application should verify that the non-compound RTCP packets are received successfully by the other end. Verification can be done by implicit means, for instance the RTCP sender that sends a feedback requests is expected to see some kind of a response to the requests in the media stream. If verification fails the RTCP sender shall switch to the use of compound RTCP packets according to the rules outlined in RFC3550 [9].

7.4 RTP payload formats for terminals

7.4.1 General

This clause specifies RTP payload formats for terminals for all codecs supported by MTSI in clause 5.2. Note that each RTP payload format also specifies media type signalling for usage in SDP.

7.4.2 Speech

When transmitting AMR or AMR-WB encoded media in RTP

- the AMR (and AMR-WB) payload format shall be used [28].

MTSI terminals shall support both the bandwidth-efficient and the octet-aligned payload format. The bandwidth-efficient payload format shall be preferred over the octet-aligned payload format.

The MTSI terminal shall use the parameters defined in table 7.1 during the session, unless the remote side prevents it. For all access technologies, and for normal operating conditions, the MTSI terminal should encapsulate the number of non-redundant (a.k.a. primary) speech frames in the RTP packets that corresponds to the ptime value defined in table 7.1. The MTSI terminal may encapsulate more non-redundant speech frames in the RTP packet but shall not encapsulate more than 4 non-redundant speech frames in the RTP packets. The MTSI terminal may encapsulate any number of redundant speech frames in an RTP packet but the length of an RTP packet, measured in ms, shall never exceed the maxptime value.

NOTE: The terminology "non-redundant speech frames" refers to speech frames that have not been transmitted in any preceding packet.

Table 7.1: Encapsulation parameters (to be used as defined above)

Radio access bearer technology	Recommended encapsulation	ptime	maxptime
Unknown	1 non-redundant speech frame per RTP packet Max 12 speech frames in total	20	240
HSPA	1 non-redundant speech frame per RTP packet Max 12 speech frames in total	20	240
EGPRS	2 non-redundant speech frames per RTP packet Max 12 speech frames in total	40	240
GIP	1 to 12 non-redundant speech frames per RTP packet Max 12 speech frames in total	20, 40, 60 or 80	240

NOTE: It is possible to send only redundant speech frames in one RTP packet.

For all radio access bearer technologies, the bandwidth-efficient payload format should be used unless the session setup concludes that the octet-aligned payload format is the only payload format that all parties support. The SDP offer shall include an RTP payload type where octet-align=0 is defined or where octet-align is not specified and should include another RTP payload type with octet-align=1. MTSI terminals offering wide-band speech shall offer these parameters and parameter settings also for the RTP payload types used for wide-band speech.

For examples of SDP offers and answers, see annex A.

The RTP payload format for DTMF events ís described in Annex G.

7.4.3 Video

The following RTP payload formats shall be used:

- H.263 video codec RTP payload format according to RFC 4629 [29];
- MPEG-4 video codec RTP payload format according to RFC 3016 [30];

- H.264 (AVC) video codec RTP payload format according to RFC 3984 [25], where the interleaved packetization mode shall not be used. Receivers shall support both the single NAL unit packetization mode and the non-interleaved packetization mode of RFC 3984 [25], and transmitters may use either one of these packetization modes.

7.4.4 Real-time text

The following RTP payload format shall be used:

- T.140 text conversation RTP payload format according to RFC 4103 [31].

Real-time text shall be the only payload type in its RTP stream because the RTP sequence numbers are used for loss detection and recovery. The redundant transmission format shall be used for keeping the effect of packet loss low.

Media type signalling for usage in SDP is specified in section 10 of RFC 4103 [31] and section 3 of RFC 4102 [49].

7.5 Media flow

7.5.1 General

This clause contains considerations on how to use media in RTP, packetization guidelines, and other transport considerations.

7.5.2 Media specific

7.5.2.1 Speech

7.5.2.1.1 General

This clause describes how the voice media should be packetized during a session. It includes definitions both for the cases where the access type is known and one default operation for the case when the access type is not known.

Requirements for transmission of DTMF events are described in Annex G.

7.5.2.1.2 Default operation

If AMR is used, the codec mode set Config-NB-Code=1 [16] {AMR-NB12.2, AMR-NB7.4, AMR-NB5.9 and AMR-NB4.75} should be used unless the session-setup negotiation determines that other codec modes shall be used.

If AMR-WB is used, the codec mode set Config-WB-Code=0 [16] {AMR-WB12.65, AMR-WB8.85 and AMR-WB6.60} should be used unless the session-setup negotiation determines that other codec modes shall be used.

In the transmitted media, codec mode changes should be aligned to every other frame border and should be performed to one of the neighbouring codec modes in the negotiated mode set. In the received media, codec mode changes shall be accepted at any frame border and to any codec mode within the negotiated mode set.

The adaptation of codec mode, aggregation and redundancy is defined in clause 10.2. The MTSI terminal should not set the CMR bits in the AMR payload format. It shall however accept requests signalled with the CMR bits.

The AMR bandwidth-efficient payload format should be used unless the session setup determines that the octet-aligned payload format must be used.

The terminal should send one speech frame encapsulated in each RTP packet unless the session setup defines that the other PS end-point wants to receive another encapsulation variant.

The terminal should request to receive one speech frame encapsulated in each RTP packet but shall accept any number of frames per RTP packet up to the maximum limit of 12 speech frames per RTP packet.

For application-layer redundancy, see clause 9.2.

7.5.2.1.3 HSPA

Use default operation.

NOTE: The RLC PDU sizes have been optimized for the codec modes, payload formats and frame encapsulations defined in the default operation in clause 7.5.2.1.2 of 3GPP 26.131 [35].

7.5.2.1.4 EGPRS

Use default operation, except that the terminal

- should send two speech frames encapsulated in each RTP packet;
- should request receiving two speech frames encapsulated in each RTP packet unless the session setup defines that other PS end-point want to receive another encapsulation variant.

7.5.2.1.5 GIP

Use default operation, except that the terminal:

- should send 0, 1, 2, 3 or 4 non-redundant speech frames encapsulated in each RTP packet and should request receiving 1 to 4 speech frames in each RTP packet;
- may use application layer redundancy, in which case the terminal may encapsulate up to 12 speech frames in each RTP packet, with a maximum of four non-redundant speech frames and maximum 8 redundant speech frames.

7.5.2.2 Video

An MTSI terminal should follow general strategies for error-resilient coding (segmentation) and packetization as specified by each codec [22], [23], [24] and RTP payload format [25], [29], [39] specification. Further guidelines on how the video media data should be packetized during a session are provided in this clause.

Coded pictures should be encoded into individual segments:

- For H.263 Profile 0, a Picture Start Code (PSC) or non-empty Group of Block (GOB) header indicates the beginning of such a segment.
- For H.263 Profile 3, MPEG-4 (Part 2) Visual, and H.264 / MPEG-4 (Part 10) AVC, a slice corresponds to such a segment.

Each individual segment should be encapsulated in one RTP packet. Each RTP packet should be smaller than the Maximum Transfer Unit (MTU) size.

NOTE 1: Unnecessary video segmentation, e.g. within RTP packets, may reduce coding efficiency.

NOTE 2: RTP packet fragmentation, e.g. across UDP boundaries, may decrease transport overhead and reduce error robustness. Hence, packet size granularity is a trade-off between error robustness and overhead that may be tuned according to bearer access characteristics if available.

NOTE 3: In most cases, the MTU-size has a direct relationship with the bearer of the radio network.

7.5.2.3 Text

Real-time text is intended for human conversation applications. Text shall not be transferred with higher rate than 30 characters per second (as defined for cps in section 6 of RFC 4103 [31]). A text-capable MTSI terminal shall be able to receive text with cps set up to 30.

7.5.3 Media synchronization

7.5.3.1 General

RTCP SR shall be used for media synchronization by setting the NTP and RTP timestamps according to RFC 3550 [9]. To enable quick media synchronization when a new media component is added, or an MTSI session is initiated, the RTP sender should send RTCP Sender Reports for all newly started media components as early as possible.

NOTE: An MTSI sender can signal in SDP that no synchronization between media components is required. See clause 6.2.6 and clause A.7.

7.5.3.2 Text

The media synchronization requirements for real-time text are relaxed. A synchronization error between text and other media of a maximum of 3 seconds is accepted. Since this is longer than the maximum accepted latency, no specific methods need to be applied to assure to meet the requirement

8 Jitter buffer management in terminals

8.1 General

This clause specifies mechanisms to handle delay jitter in terminals.

8.2 Speech

8.2.1 Terminology

In the following paragraph(s), Jitter Buffer Management (JBM) denotes the actual buffer as well as any control, adaptation and media processing algorithm (excluding speech decoder) used in the management of the jitter induced in the transport channel. An illustration of an exemplary structure of an MTSI speech receiver with adaptive jitter buffer is shown in figure 8.1 to clarify the terminology and the relation between different functional components.

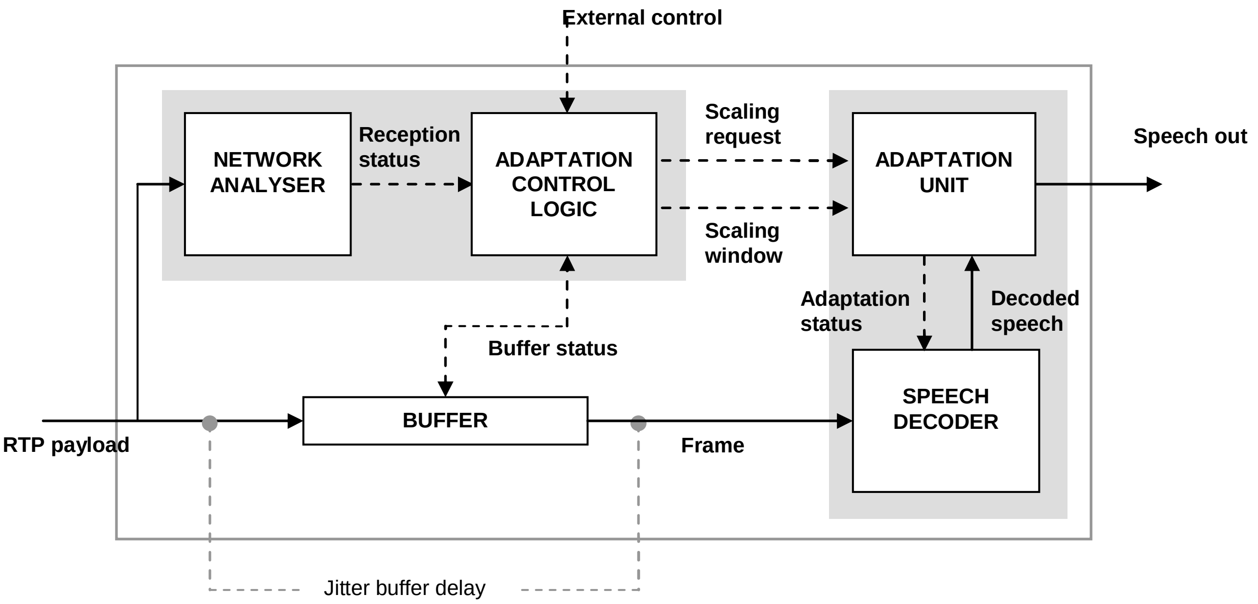


Figure 8.1: Example structure of an MTSI speech receiver

The blocks "network analyzer" and "adaptation control logic" together with the information on buffer status form the actual buffer control functionality, whereas "speech decoder" and "adaptation unit" provide the media processing functionality. Note that the external playback device control driving the media processing is not shown in figure 8.1.

The grey dashed lines indicate the measurement points for the jitter buffer delay, i.e. the difference between the decoder consumption time and the arrival time of the speech frame to the JBM.

The functional processing blocks are as follows:

- **Buffer:** The jitter buffer unpacks the incoming RTP payloads and stores the received speech frames. The buffer status may be used as input to the adaptation decision logic. Furthermore, the buffer is also linked to the speech decoder to provide frames for decoding when they are requested for decoding.

- **Network analyser:** The network analysis functionality is used to monitor the incoming packet stream and to collect reception statistics (e.g. jitter, packet loss) that are needed for jitter buffer adaptation. Note that this block can also include e.g. the functionality needed to maintain statistics required by the RTCP if it is being used.

- **Adaptation control logic:** The control logic adjusting playback delay and operating the adaptation functionality makes decisions on the buffering delay adjustments and required media adaptation actions based on the buffer status (e.g. average buffering delay, buffer occupancy, etc.) and input from the network analyser. Furthermore, external control input can be used e.g. to enable inter-media synchronisation or other external scaling requests. The control logic may utilize different adaptation strategies such as fixed jitter buffer (without adaptation and time scaling), simple adaptation during comfort noise periods or buffer adaptation also during active speech. The general operation is controlled with desired proportion of frames arriving late, adaptation strategy and adaptation rate.

- **Speech decoder:** The standard AMR or AMR-WB speech decoder. Note that the speech decoder is also assumed to include error concealment / bad frame handling functionality. Speech decoder may be used with or without the adaptation unit.

- **Adaptation unit:** The adaptation unit shortens or extends the output signal length according to requests given by the adaptation control logic to enable buffer delay adjustment in a transparent manner. The adaptation is performed using the frame based or sample based time scaling on the decoder output signal during comfort noise periods only or during active speech and comfort noise. The buffer control logic should have a mechanism to limit the maximum scaling ratio. Providing a scaling window in which the targeted time scale modifications are performed improves the situation in certain scenarios - e.g. when reacting to the clock drift or to a request of inter-media (re)synchronization - by allowing flexibility in allocating the scaling request on several frames and performing the scaling on a content-aware manner. The adaptation unit may be implemented either in a separate entity from the speech decoder or embedded within the decoder.

8.2.2 Functional requirements for jitter-buffer management

The functional requirements for the speech JBM guarantee appropriate management of jitter which shall be the same for all speech JBM implementations used in MTSI clients. A JBM implementation used in MTSI shall support the following requirements, but is not limited in functionality to these requirements. They are to be seen as a minimum set of functional requirements supported by every speech JBM used in MTSI.

Speech JBM used in MTSI shall:

- support all the codecs as defined in clause 5.2.1;

- support source-controlled rate operation as well as non-source-controlled rate operation;

- be able to receive the de-packetized frames out of order and present them in order for decoder consumption;

- be able to receive duplicate speech frames and only present unique speech frames for decoder consumption;

- be able to handle clock drift between the encoding and decoding end-points.

8.2.3 Minimum performance requirements for jitter-buffer management

8.2.3.1 General

The jitter buffering time is the time spent by a speech frame in the JBM. It is measured as the difference between the decoding start time and the arrival time of the speech frame to the JBM. The frames that are discarded by the JBM are not counted in the measure.

The minimum performance requirements consist of objective criteria for delay and jitter-induced concealment operations. In order for a JBM implementation to pass the minimum performance requirements all objective criteria shall be met.

A JBM implementation used in MTSI shall comply with the following design guidelines:

1. The overall design of the JBM shall be to minimize the buffering time at all times while still conforming to the minimum performance requirements of jitter induced concealment operations and the design guidelines for sample-based timescaling (as set in bullet point 3);
2. If the limit of jitter induced concealment operations cannot be met, it is always preferred to increase the buffering time in order to avoid growing jitter induced concealment operations going beyond the stated limit above. This guideline applies even if that means that end-to-end delay requirement given in 3GPP TS 22.105 [34] can no longer be met;
3. If sample-based time scaling is used (after speech decoder), then artefacts caused by time scaling operation shall be kept to a minimum. Time scaling means the modification of the signal by stretching and/or compressing it over the time axis. The following guidelines on time scaling apply:
   - Use of a high-quality time scaling algorithm is recommended;
   - The amount of scaling should be as low as possible;
   - Scaling should be applied as infrequently as possible;
   - Oscillating behaviour is not allowed.

NOTE: If the end-to-end delay for the ongoing session is known to the MTSI client and measured to be less than 150 ms (as defined in 3GPP TS 22.105 [34]), the JBM may relax its buffering time minimization criteria in favour of reduced JBM adaptation artefacts if such a relaxation will improve the media quality. Note that a relaxation is not allowed when testing for compliance with the minimum performance requirements specified in clauses 8.2.3.2.2 and 8.2.3.2.3.

8.2.3.2 Objective performance requirements

8.2.3.2.1 General

The objective performance requirements consist of criteria for delay, time scaling and jitter-induced concealment operations.

The objective minimum performance requirements are divided into three parts:

1. Limiting the jitter buffering time to provide as low end-to-end delay as possible.
2. Limiting the jitter induced concealment operations, i.e. setting limits on the allowed induced losses in the jitter buffer due to late losses, re-bufferings, and buffer overflows.
3. Limiting the use of time scaling to adapt the buffering depth in order to avoid introducing time scaling artefacts on the speech media.

In order to fulfil the objective performance requirements, the JBM under test needs to pass the respective criteria using the six channels as defined in clause 8.2.3.3. Note that in order to pass the criteria for a specific channel, all three requirements must be fulfilled.

8.2.3.2.2 Jitter buffer delay criteria

The reference delay computation algorithm in Annex D defines the performance requirements for the set of delay and error profiles described in clause 8.2.3.3. The JBM algorithm under test shall meet these performance requirements. The performance requirements shall be a threshold for the Cumulative Distribution Function (CDF) of the speech-frame delay introduced by the reference delay computation algorithm. A CDF threshold is set by shifting the reference delay computation algorithm CDF 60 ms. The speech-frame delay CDF is defined as:

$$P(x) = Probability\ (delay\_compensation\_by\_JBM \leq x)$$

The relation between the reference delay computation algorithm and the CDF threshold is outlined in figure 8.2.

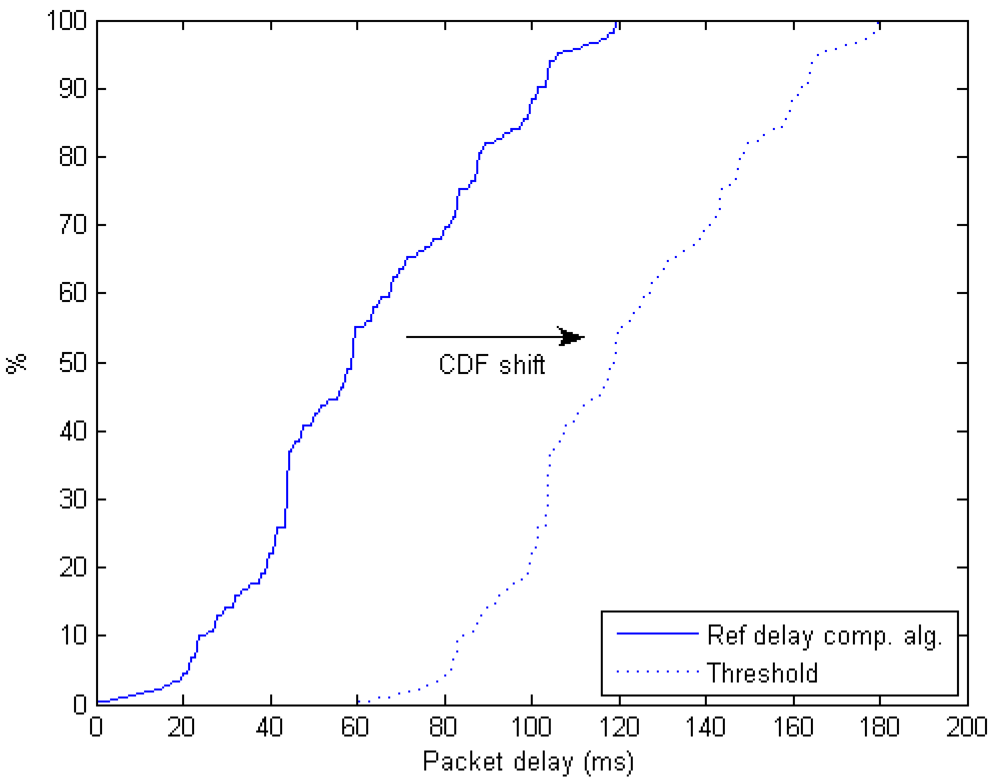


Figure 8.2: Example showing the relation between the reference delay algorithm and the CDF threshold - the delay and error profile 4 in table 8.1 has been used

The JBM algorithm under test shall achieve lower or same delay than that set by the CDF threshold for at least 90 % of the speech frames. The values for the CDF shall be collected for the full length of each delay and error profile. The delay measure in the criteria is measured as the time each speech frame spends in the JBM; i.e. the difference between the decoder consumption time and the arrival time of the speech frame to the JBM.

The parameter settings for the reference delay computation algorithm are:

- adaptation_lookback = 200;
- delay_delta_max = 20;
- target_loss= 0.5.

8.2.3.2.3 Jitter induced concealment operations

The jitter induced concealment operations include:

- JBM induced removal of a speech frame, i.e. buffer overflow or intentional frame dropping when reducing the buffer depth during adaptation.
- Deletion of a speech frame because it arrived at the JBM too late.
- Modification of the output timeline due to link loss.
- Jitter-induced insertion of a speech frame controlled by the JBM (e.g. buffer underflow).

Link losses handled as error concealment and not changing the output timeline shall not be counted in the jitter induced concealment operations.

Jitter loss rate = JBM triggered concealed frames / Number of transmitted frames

The jitter loss rate shall be calculated for active speech frames only.

NOTE: SID_FIRST and SID_UPDATE frames belong to the non-active speech period, hence concealment for losses of such frames should not be included in the statistics.

The jitter loss rate shall be below 1% for every channel measured over the full length of the respective channel. The value of 1 % was chosen because such a loss rate will usually not significantly reduce the speech quality.

8.2.3.3 Delay and error profiles

Six different delay and error profiles are used to check the tested JBM for compliance with the minimum performance requirements. The profiles span a large range of operating conditions in which the JBM shall provide sufficient performance for the MTSI service. All profiles are 7 500 IP packets long.

Table 8.1: Delay and error profile overview - The channels are attached electronically

Profile	Characteristics	Packet loss rate (%)	Filename
1	Low-amplitude, static jitter characteristics, 1 frame/packet	0	dly_error_profile_1.dat
2	Hi-amplitude, semi-static jitter characteristics, 1 frame/packet	0.24	dly_error_profile_2.dat
3	Low/high/low amplitude, changing jitter, 1 frame/packet	0.51	dly_error_profile_3.dat
4	Low/high/low/high, changing jitter, 1 frame/packet	2.4	dly_error_profile_4.dat
5	Moderate jitter with occasional delay spikes, 2 frames/packet (7 500 IP packets, 15 000 speech frames)	5.9	dly_error_profile_5.dat
6	Moderate jitter with severe delay spikes, 1 frame/packet	0.1	dly_error_profile_6.dat

The attached profiles in the zip-archive "delay_and_error_profiles.zip" are formatted as raw text files with one delay entry per line. The delay entries are written in milliseconds and packet losses are entered as "-1". Note that when testing for compliance, the starting point in the delay and error profile shall be randomized.

8.2.3.4 Speech material for JBM minimum performance evaluation

The files described in table 8.2 and attached to the present document in the zip-archive "JBM_evaluation_files.zip" shall be used for evaluation of a JBM against the minimum performance requirements. The data is stored as RTP packets, formatted according to "RTP dump" format [41]. The input to these files is AMR or AMR-WB encoded frames, encapsulated into RTP packets using the octet-aligned mode of the AMR RTP payload format [28].

Table 8.2: Input files for JBM performance evaluation - The files are attached electronically

Codec	Frames per RTP packet	Filename
AMR (12.2 kbps)	1	test_amr122_fpp1.rtp
AMR (12.2 kbps)	2	test_amr122_fpp2.rtp
AMR-WB (12.65 kbps)	1	test_amrwb1265_fpp1.rtp
AMR-WB (12.65 kbps)	2	test_amrwb1265_fpp2.rtp

8.3 Video

Video receivers should implement an adaptive video de-jitter buffer. The overall design of the buffer should aim to minimize delay, maintain synchronization with speech, and minimize dropping of late packets. The exact implementation is left to the implementer.

8.4 Text

Conversational quality of real-time text is experienced as being good, even with up to one second end-to-end text delay. Strict jitter buffer management is therefore not needed for text. Basic jitter buffer management for text is described in section 5 of RFC 4103 [31] where a calculation is described for the time allowed before an extra delayed text packet may be regarded to be lost.

9 Packet-loss handling

9.1 General

This clause specifies some methods to handle conditions with packet losses. Packet losses in general will also trigger adaptation, which is specified in clause 10.

9.2 Speech

9.2.1 General

This clause provides a recommendation for a simple application layer redundancy scheme that is useful in order to handle operational conditions with severe packet loss rates. Simple application layer redundancy is generated by encapsulating one or more previously transmitted speech frames into the same RTP packet as the current previously not transmitted frame(s). An RTP packet may thus contain zero, one or several redundant speech frames and zero, one or several non-redundant speech frames.

When transmitting redundancy, the terminal should switch to a lower codec mode. The terminal shall utilize the codec mode rates within the negotiated codec mode set with the negotiated adaptation steps and limitations as defined by mode-change-neighbor and mode-change-period. It is recommended to not send redundant speech frames before the targeted codec mode is reached. Table 9.1 defines the recommended codec modes for different redundancy level combinations.

When application layer redundancy is used for AMR or AMR-WB encoded speech media, the transmitting application may use up to 300 % redundancy, i.e. a speech frame transported in one RTP packet may be repeated in 3 other RTP packets.

Table 9.1: Recommended codec modes and redundancy level combinations when redundancy is supported

Redundancy level	No redundancy	100 % redundancy
Narrow-band speech	AMR 12.2	AMR 5.9
Wide-band speech (when wide-band is supported)	AMR12.65	AMR 6.60

9.2.2 Transmitting redundant frames

When transmitting redundant frames, the redundant frames should be encapsulated together with non-redundant media data as shown in figure 9.1. The frames shall be consecutive with the oldest frame placed first in the packet and the most recent frame placed last in the packet. The RTP Timestamp shall represent the sampling time of the first sample in the oldest frame transmitted in the packet.

NOTE: When switching from no redundancy to using redundancy, the RTP Timestamp may be the same for consecutive RTP packets.

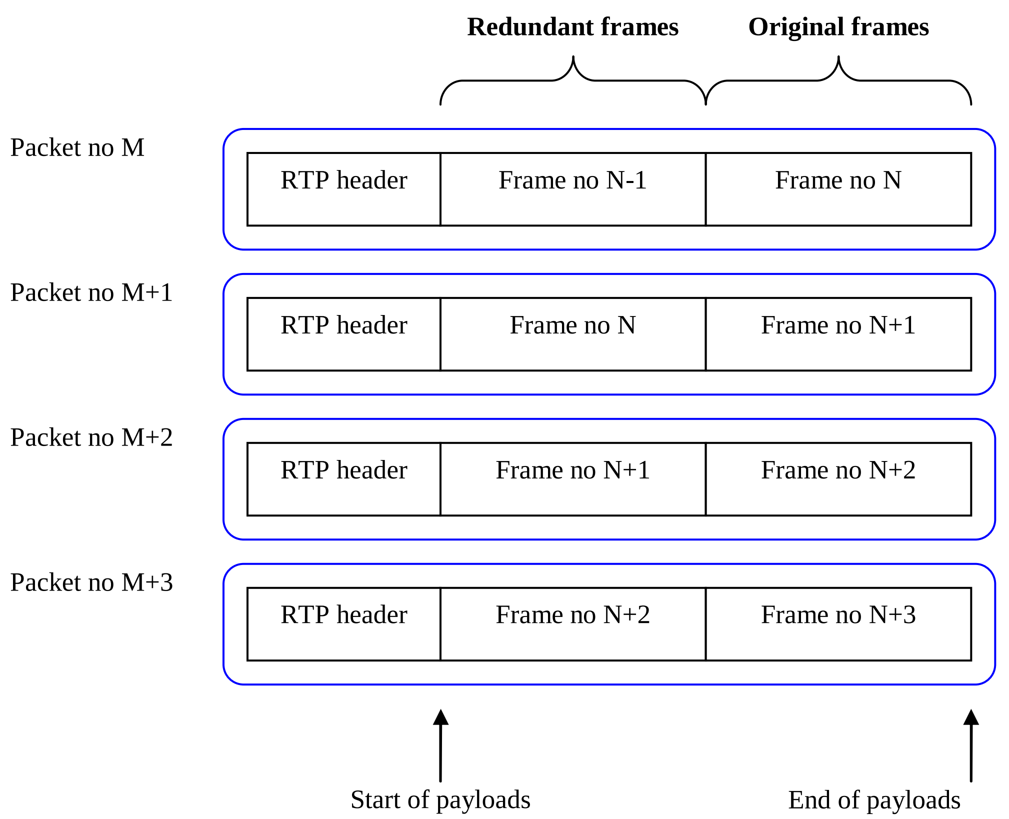


Figure 9.1: Redundant and non-redundant frames in the case of 100 % redundancy, when the original packing is 1 frame per packet

Figure 9.1 shows only one non-redundant frame encapsulated together with one redundant frame. It is allowed to encapsulate several non-redundant frames with one or several redundant frames. The following combinations of non-redundant frames and redundant frames can be used.

Table 9.2: Example frame encapsulation with different redundancy levels and when maxptime is 240

Original encapsulation (without redundancy)	Encapsulation with 100 % redundancy	Encapsulation with 200 % redundancy	Encapsulation with 300 % redundancy
1 frame per packet	≤ 1 non-redundant frame and ≤ 1 redundant frame	≤ 1 non-redundant frame and ≤ 2 redundant frames	≤ 1 non-redundant frame and ≤ 3 redundant frames
2 frames per packet	≤ 2 non-redundant frames and ≤ 2 redundant frames	≤ 2 non-redundant frames and ≤ 4 redundant frames	≤ 2 non-redundant frames and ≤ 6 redundant frames
3 frames per packet	≤ 3 non-redundant frames and ≤ 3 redundant frames	≤ 3 non-redundant frames and ≤ 6 redundant frames	≤ 3 non-redundant frames and ≤ 9 redundant frames
4 frames per packet	≤ 4 non-redundant frames and ≤ 4 redundant frames	≤ 4 non-redundant frames and ≤ 8 redundant frames	Not allowed since maxptime does not allow more than 12 frames per RTP packet in this example

With a maxptime value of 240, it is possible to encapsulate up to 12 frames per packet. It is therefore not allowed to use 300 % when the original encapsulation is 4 frames per packet, as shown in table 9.2. If the receiver's maxptime value is lower than 240 then even more combinations of original encapsulation and redundancy level will be prohibited.

Figure 9.2 shows an example where the frame aggregation is 2 frames per packet and when 100 % redundancy added.

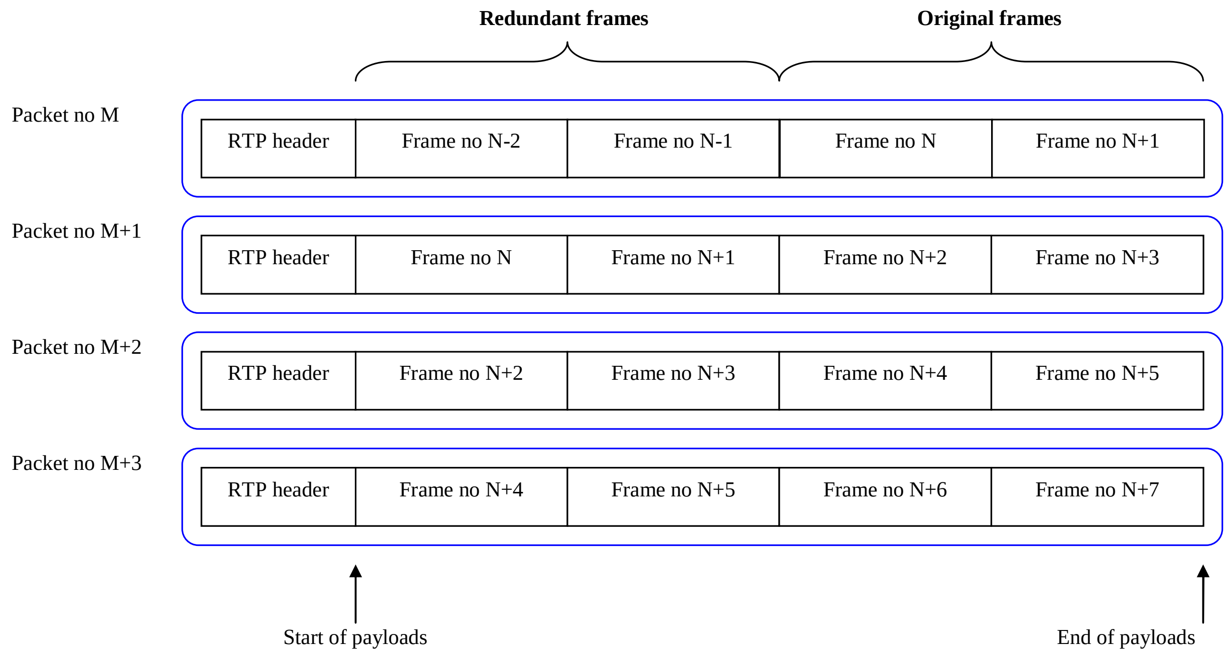


Figure 9.2: Redundant and non-redundant frames in the case of 100 % redundancy, when the original packing is 2 frames per packet

A redundant frame may be replaced by a NO_DATA frame. If the transmitter wants to encapsulate non-consecutive frames into one RTP packet, then NO_DATA frames shall be inserted for the frames that are not transmitted in order to create frames that are consecutive within the packet. This method is used when sending redundancy with an offset, see figure 9.3.

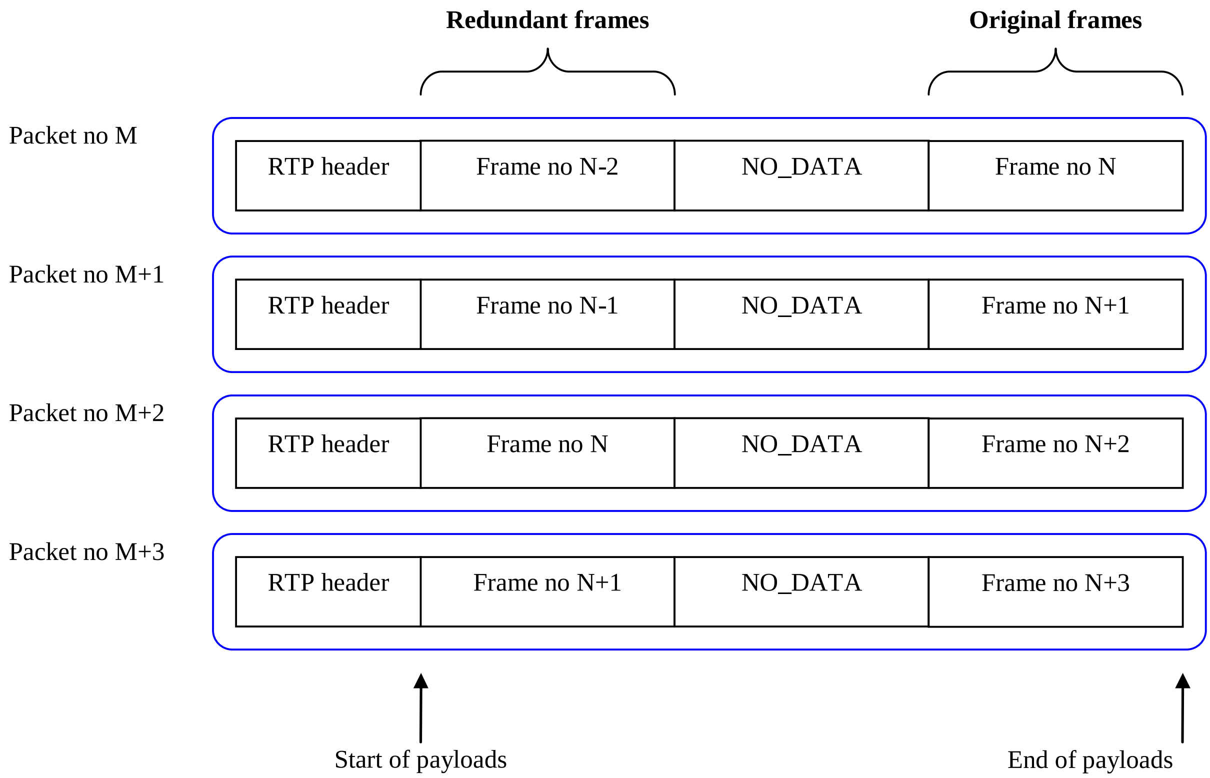


Figure 9.3: Redundant and non-redundant frames in the case of 100 % redundancy, when the original packing is 1 frame per packet and when the redundancy is transmitted with an offset of 20 ms

Note that with this scheme, the receiver may receive a frame 3 times: first the non-redundant encoding; then as a NO_DATA frame; and finally the redundant frame. Other combinations of redundancy and offset may result in receiving even more copies of a frame. The proper receiver behaviour is described in the AMR payload format [28].

For any combinations of frame aggregation, redundancy and redundancy offset, the transmitter shall not exceed the frame encapsulation limit indicated by the receiver's maxptime value when constructing the RTP packet.

When source controlled rate operation is used, it is allowed to send redundant media data without any non-redundant media, if no non-redundant media is available.

NOTE 1: When going from active speech to DTX, there may be no non-redundant frames in the end of the talk spurt while there still are redundant frames that need to be transmitted.

In the end of a talk spurt, when there are no more non-redundant frames to transmit, it is allowed to drop the redundant frames that are in the queue for transmission.

NOTE 2: This ensures that it is possible to use redundancy without increasing the packet rate. The quality degradation by having less redundancy for the last frames should be negligible since these last frames typically contain only background noise.

NOTE 3: The RTP Marker Bit shall be set according to Section 4.1 of the AMR payload format [28].

9.2.3 Receiving redundant frames

In order to receive and decode redundant media properly, the receiving application shall sort the received frames based on the RTP Timestamp and shall remove duplicated frames. If multiple versions of a frame are received, i.e. encoded with different bitrates, then the frame encoded with the highest bitrate should be used for decoding.

9.3 Video

AVPF NACK messages are used by MTSI terminals to indicate non-received RTP packets for video (see clause 7.3.3). An MTSI terminal transmitting video can use this information, as well as the AVPF Picture Loss Indication (PLI), to accommodate for losses in the encoding process. See also clause 7.5.2.2 on error-resilient video coding,

9.4 Text

Redundant transmission provided by the RTP payload format as described in RFC 4103 [31] shall be supported. The transmitting application may use up to 200 % redundancy, i.e. a T140block transported in one RTP packet may be repeated once or twice in subsequent RTP packets. 200 % redundancy shall be used when the conditions along the call path are not known to be free of loss. However, the result of media negotiation shall be followed, and transmission without redundancy used if one of the parties does not show capability for redundancy.

The sampling time shall be 300 ms as a minimum (in order to keep the bandwidth down) and should not be longer than 500 ms. New text after an idle period shall be sent as soon as possible. The first packet after an idle-period shall have the M-bit set.

The procedure described in section 5 of RFC 4103 [31], or a procedure with equivalent or better performance, shall be used for packet-loss handling in the receiving terminal.

10 Adaptation

10.1 General

Adaptive mechanisms are used to optimize the session quality given the current transport characteristics. The mechanisms provided in MTSI are bit-rate, packet-rate and error resilience adaptation. These mechanisms can be used in different ways; however, they should only be used when the result of the adaptation is assumed to increase the session quality even if e.g. the source bit-rate is reduced.

Adaptive mechanisms that act upon measured or signalled changes in the transport channel characteristics may be used in a conservative manner. A conservative use of adaptation is characterized by a fast response to degrading conditions, and a slower, careful upwards adaptation intended to return the session media settings to the original default state of the session. The long-term goal of any adaptive mechanism is assumed to be a restoration of the session quality to the originally negotiated quality. The short-term goal is to maximize the session quality given the current transport characteristics, even if that means than the adapted state of the session will give a lower session quality compared to the session default state if transported on an undisturbed channel.

10.2 Speech

10.2.1 RTCP-APP with codec control requests

When signalling adaptation requests for speech in MTSI, an RTCP-APP packet shall be used. This application-specific packet format supports three different adaptation requests; bit-rate requests, packet rate requests and redundancy requests. The RTCP-APP packet is put in a compound RTCP packets according to the rules outlined in RFC 3550 [9] and RFC 4585 [40]. In order to keep the size of the RTCP packets as small as possible it is strongly recommended that the RTCP packets are transmitted as minimal compound RTCP packets, meaning that they contain only the items:

- SR or RR;
- SDES CNAME item;
- APP (when applicable).

The recommended RTCP mode is RTCP-AVPF early mode since it will enable transmission of RTCP reports when needed and still comply with RTCP bandwidth rules. The RTCP-APP packets should not be transmitted in each RTCP packet, but rather as a result in the transport characteristics which require end-point adaptation.

The signalling allows for a request that the other endpoint modifies the packet stream to better fit the characteristics of the current transport link. Note that the media sender can, if having good reasons, choose to not comply with the request received from the media receiver. One such reason could be knowledge of that the local conditions do not allow the requested format.

The RTCP-APP packet defined to be used for adaptation signalling for speech in MTSI is constructed as shown in figure 10.1.

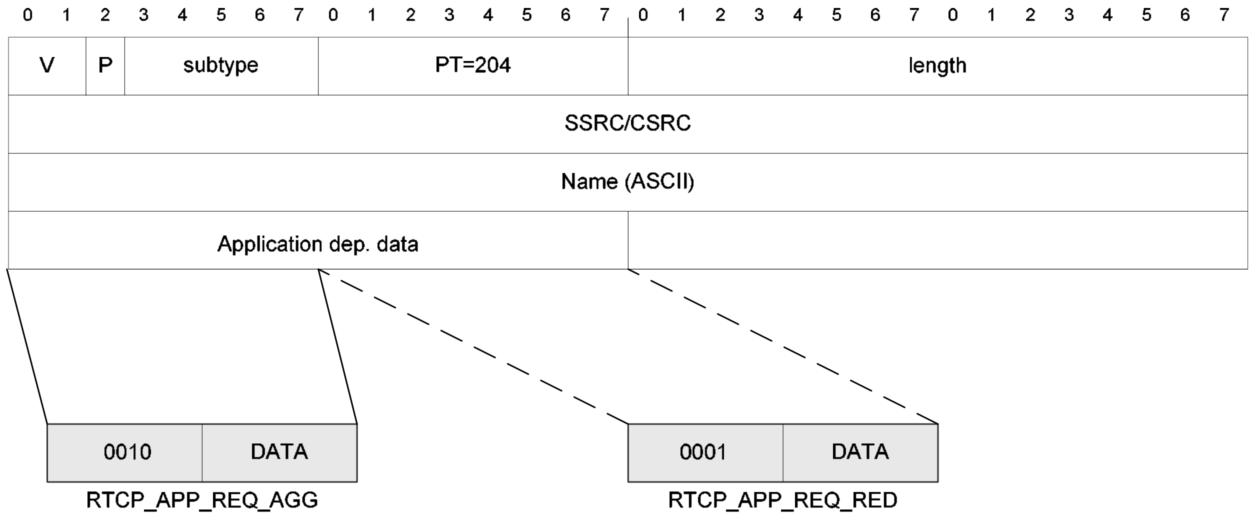


Figure 10.1: RTCP-APP formatting

The RTCP-APP specific fields are defined as follows:

- Subtype - the subtype value shall be set to "0".
- Name - the name shall be set to "3GM7", meaning 3GPP MTSI Release 7.

The application-dependent data field contains the requests listed below. The length of the application-dependent data shall be a multiple of 32 bits. The unused bytes shall be set to zero.

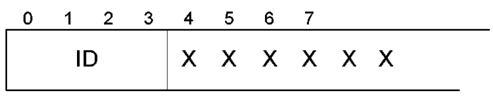


Figure 10.2: Basic syntax of the application-dependent data fields

The length of the messages is 1 or 2 bytes depending on request type.

The ID field identifies the request type. ID Code points [0000], [0001], [0010] and [0011] are specified in the present document, whereas the other ID code points are reserved for future use.

The signalling for three different is defined.

RTCP_APP_REQ_RED: Request for redundancy level and offset of redundant data.

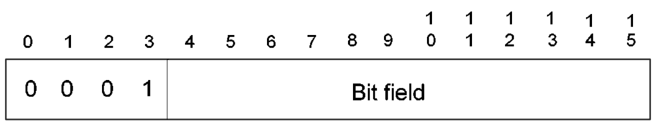


Figure 10.3: Redundancy request

The Bit field is a 12 bit bitmask that signals a request on how non-redundant payloads chunks are to be repeated in subsequent packets.

The position of the bit set indicates which earlier non-redundant payload chunks is requested to be added as redundant payload chunks to the current packet.

- If the LSB (rightmost bit) is set equal to 1 it indicates that the last previous payload chunk is requested to be repeated as redundant payload in the current packet.
- If the MSB (leftmost bit) is set equal to 1 it indicates that the payload chunk that was transmitted 12 packets ago is requested to be repeated as redundant payload chunk in the current packet. Note that it is not guaranteed that the sender has access to such old payload chunks.

The maximum amount of redundancy is 300 %, i.e., at maximum three bits can be set in the Bit field.

See clause 10.2.1 for example use cases.

RTCP_APP_REQ_AGG: Request for a change of frame aggregation.

0	1	2	3	4	5	6	7
0	0	1	0	DATA			

Figure 10.4: Frame aggregation request

The DATA field is a 4 bit value field:

- 0000 - 1 frame / packet.
- 0001 - 2 frames / packet.
- 0010 - 3 frames / packet.
- 0011 - 4 frames / packet.

The values 0100…1111 are reserved for future use.

The maximum allowed frame aggregation is also limited by the maxptime parameter in the session SDP since the sender is not allowed to send more frames in an RTP packet than what the maxptime parameter defines.

The default aggregation is governed by the ptime parameter in the session SDP. It is allowed to send fewer frames in an RTP packet, for example if there are no more frames available at the end of a talk spurt. It is also allowed to send more frames in an RTP packet, but such behaviour is not recommended.

See clauses 7.4.2 and 12.3.2.1 for further information.

RTCP_APP_CMR: Codec Mode Request

0	1	2	3	4	5	6	7
0	0	1	1	CMR			

Figure 10.5: Codec mode request

The CMR bits are identical to the CMR bits defined in [28]. For an IP endpoint the CMR should be transmitted in an RTCP_APP_CMR, the CMR bits in the AMR payload should be left unchanged.

If the session is an interworking session with a legacy circuit-switched (CS) system, the CMR bits from the CS client should be forwarded in the CMR bits in the AMR payload, the RTCP_APP_CMR should not be used in the direction from the media gateway towards the PS client. If CMR bits are available both in the payload and in an RTCP-APP message, the payload requests shall have precedence.

Figure 10.6 below illustrates how the three requests are used by the transmitter. In this case, RTCP_APP_REQ_RED is equal to "000000000101".

- The speech encoder generates frames every 20 ms.
- The speech frames are buffered until it is possible to generate a payload chunk with the number of frames requested by RTCP_APP_REQ_AGG.
- The current payload chunk is used when constructing the current RTP packet.
- The history buffer contains previously transmitted payload chunks. The length of this buffer needs to be dimensioned to store the maximum number of payload chunks that are possible. This value is based on the max-red value, the maxptime values and from the minimum number of frames that the transmitter will encapsulate in the RTP packets. In this case, the buffer length is selected to 12 payload chunks since this corresponds to the worst case of max-red=240, maxptime=240 and one frame per payload chunk.
- After transmitting the current RTP packet, the content of the history buffer is shifted, the current payload chunk is shifted in to the history buffer as P(n-1) and the oldest payload chunk P(n-12) is shifted out.
- When constructing the (provisional) RTP payload, the selected preceding payload chunks are selected from the history buffer and added to the current payload chunk. In order to form a valid RTP payload, the transmitter needs to verify that the maxptime value is not exceeded. If the provisional RTP payload is longer than what maxptime allows, then the oldest speech frames shall be removed until the length (in time) of the payload no longer violates the maxptime value.

Note also that the transmitter is not allowed to send frames that are older than the max-red value that the transmitter has indicated in the SDP.

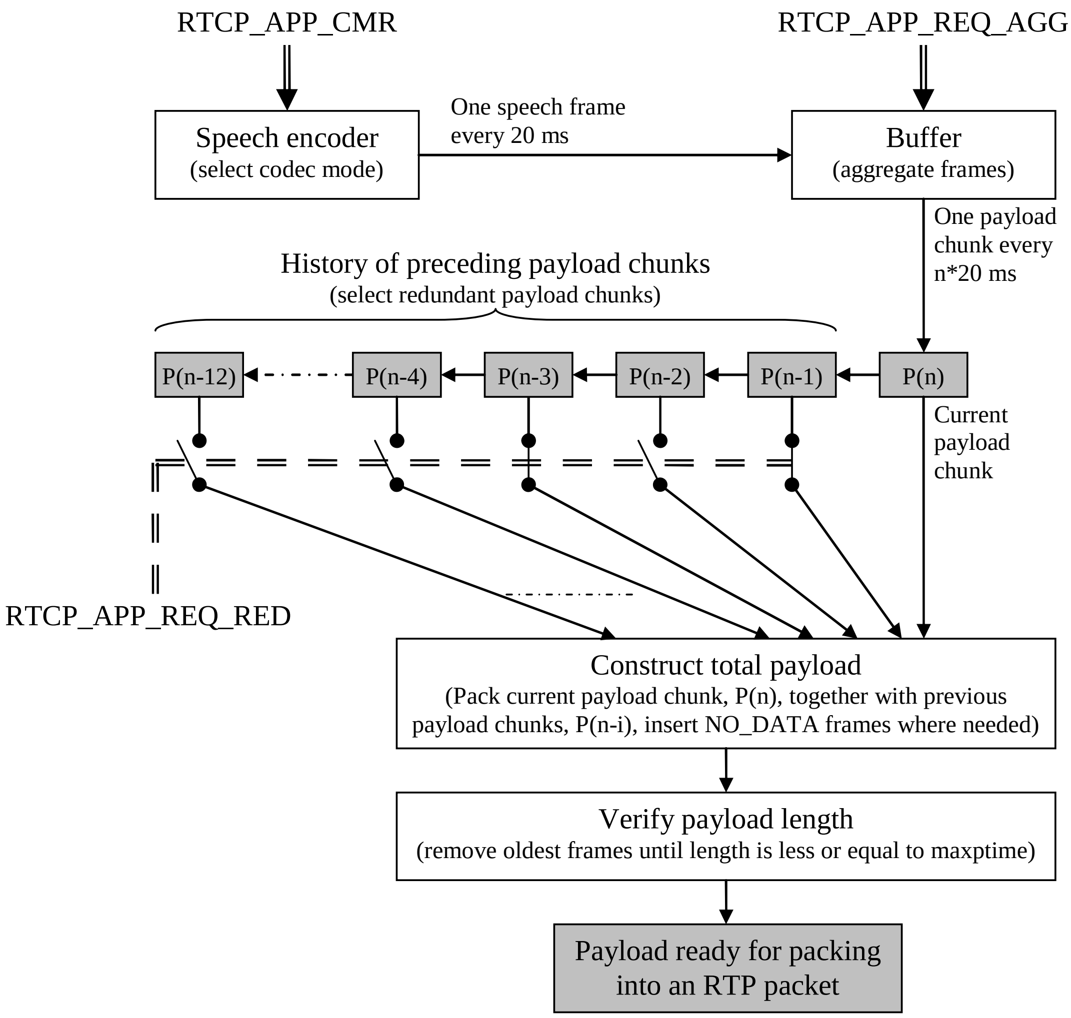


Figure 10.6: Visualization of how the different adaptation requests affect the encoding and the payload packetization

It should be noted that RTCP_APP_REQ_AGG and RTCP_APP_REQ_RED are independent. Furthermore, it should also be noted that different redundant payload chunks may contain different number of speech frames.

10.2.2 Example use cases

The following examples demonstrate how requests for redundancy and frame aggregation are realised in the RTP stream.

All examples assume that the speech codec generates frames numbered N-10…N in a continuous flow.

N-10	N-9	N-8	N-7	N-6	N-5	N-4	N-3	N-2	N-1	N

Figure 10.7: Flow of parameter sets for encoded frames Each increment corresponds to a time difference of 20 ms

In the examples below, P-1…P denote the sequence numbers of the packets.

EXAMPLE 1:

An RTCP_APP_REQ_RED request with bit field 000000000000 (no redundancy) and RTCP_APP_REQ_AGG request with value = 0 (no frame aggregation) will yield packets as shown in figure 10.8.

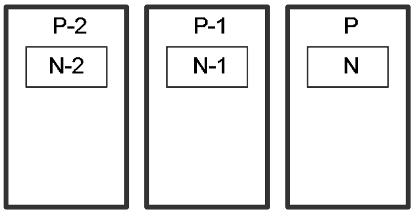


Figure 10.8: Default frame aggregation with one frame per packet

EXAMPLE 2:

An RTCP_APP_REQ_RED request with bit field 000000000001 (100% redundancy and no offset) and an RTCP_APP_REQ_AGG request with value = 0 (no frame aggregation) will yield packets as shown in figure 10.9.

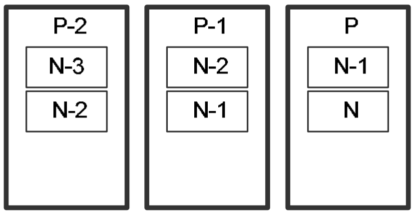


Figure 10.9: Payload packetization with 100 % redundancy and an offset of one packet

EXAMPLE 3:

An RTCP_APP_REQ_RED request with bit field 000000000010 (100% redundancy with offset 1 extra packet) and an RTCP_APP_REQ_AGG request with value = 0 (no frame aggregation) will yield packets as shown in figure 10.10.

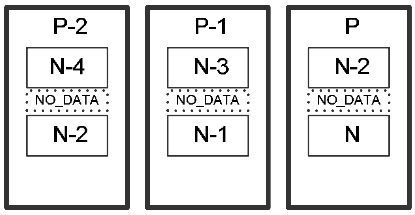


Figure 10.10: Payload packetization with 100 % redundancy and an extra offset of one packet

NO_DATA frames must be inserted to fill the gaps between two non-consecutive frames, e.g. between N-2 and N.

EXAMPLE 4:

An RTCP_APP_REQ_RED request with bit field 000000000000 (no redundancy) and RTCP_APP_REQ_AGG request with value = 1 (frame aggregation 2 frames/packet) will yield packets as shown in figure 10.11.

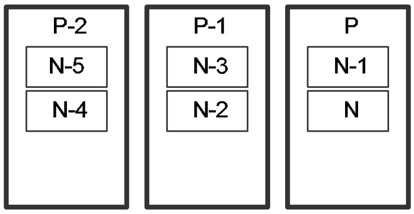


Figure 10.11: Payload packetization with 2 frames aggregated per packet

EXAMPLE 5:

An RTCP_APP_REQ_RED request with bit field 000000000001 (100% redundancy) and an RTCP_APP_REQ_AGG request with value = 1 (frame aggregation 2 frames/packet) will yield packets as shown in figure 10.12.

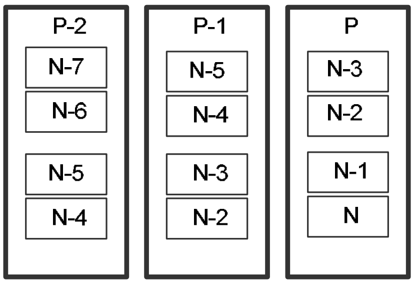


Figure 10.12: Payload packetization with 100 % redundancy and 2 frames aggregated per packet

EXAMPLE 6:

An RTCP_APP_REQ_RED request with bit field 000000000010 (100% redundancy with offset 1 extra packet) and an RTCP_APP_REQ_AGG request with value = 1 (frame aggregation 2 frames/packet) will yield packets as shown in figure 10.13.

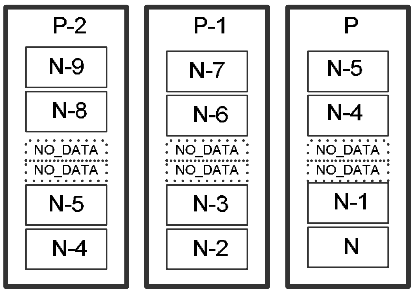


Figure 10.13: Payload packetization with 100 % redundancy, one extra offset and 2 frames aggregated per packet

10.3 Video

MTSI clients receiving RTCP Receiver Reports (RR) indicating nonzero packet loss should adjust their outgoing bitrate accordingly (see RFC 3550 [9]). Note that for IMS networks, which normally have nonzero packet loss and fairly long round-trip delay, the amount of bitrate reduction specified in RFC 3448 [56] is generally too restrictive for video and may, if used as specified, result in very low video bitrates already at (for IMS) moderate packet loss rates.

It is recommended that a video sender adapts its video output rate based on RTCP reports and TMMBR messages. Some examples are given in clause B.1.

If the receiving MTSI client is made aware of a reduction in downlink bandwidth allocation through an explicit indication from the network (e.g. due to QoS renegotiation or handoff to another radio access technology) it shall notify the sender of the new current maximum bitrate using TMMBR. In this context the TMMBR message is used to quickly signal to the other party a reduction in available bitrate. The sending client, receiving TMMBR, shall respond by sending TMMBN, as described in CCM [43]. To determine TMMBR and TMMBN content, both sending and receiving clients shall use their best estimates of packet measured overhead size when measured overhead values are not available. After receiving the TMMBN the receiving MTSI client shall send a SIP UPDATE to the other party to establish the new rate as specified in clause 6.2.7.

If the receiving MTSI client is made aware of an increase in downlink bandwidth allocation (determined via separate negotiation) through an explicit indication from the network (e.g. due to QoS renegotiation or handoff to another radio access technology) then, if this has not yet occurred, it shall send a SIP UPDATE to the other party to establish the new rate as specified in clause 6.2.7.

10.4 Text

Rate adaptation (downgrade of used bandwidth) of text shall follow the recommendation in clause 9 of RFC 4103 [31]. RTCP reports are used as indicator of loss rate over the channel.

When the transmission interval has been increased in order to handle a congestion situation, return to normal interval shall be done when RTCP reports low loss.

11 Front-end handling

11.1 General

MTSI terminals shall conform to the minimum performance requirements on the acoustic characteristics of 3G terminals specified in 3GPP TS 26.131 [35]. The codec modes and source control rate operation (DTX) settings shall be as specified in 3GPP TS 26.132 [36].

Furthermore, the test point (Point-of-Interconnect (POI)) specified in [35] shall be a reference MTSI terminal capable of receiving digital speech data at the send side and producing a digital output of the received signal (see figure 11.1). During the testing, the radio conditions should be error free and the jitter and packet loss in the IP transport shall be kept to a minimum.

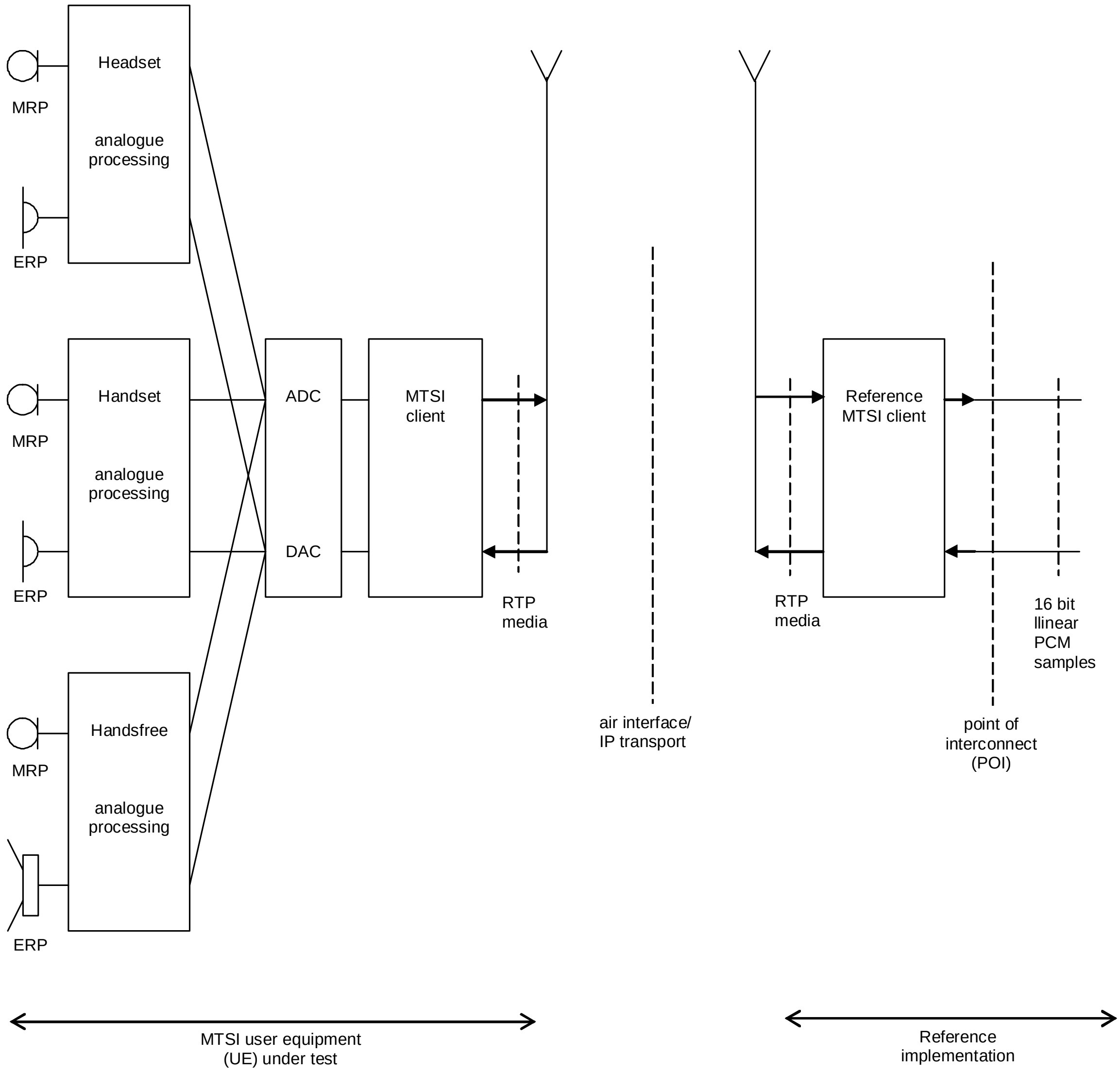


Figure 11.1: Interface for testing acoustic properties of an MTSI client

12 Inter-working

12.1 General

In order to support inter-working between different networks it is good if common codecs for the connection can be found. Requirements for different networks are described in this clause. In some cases functionality is also needed in the network to make the inter-working possible (e.g. MGCF and MGW).

NOTE: The term MGW (or Media gateway) is used in a broad sense, as it is outside the scope of the current specification to make the distinction whether certain functionality should be implemented in the MGW or in the MGCF.

12.2 3G-324M

12.2.1 General

Inter-working functions are required between IMS and CS. There are separate functions, in e.g. a MGCF, for control-plane inter-working (see 3GPP TR 29.863 [44]) and, in e.g. a MGW, for user-plane inter-working. Control-plane inter-working includes for instance SIP ⇔ BICC and SIP ⇔ H.245 protocol translations, whereas user-plane inter-working requires transport protocol translations and possibly transcoding.

12.2.2 Codec usage

12.2.2.1 General

An interoperable set of speech, video and real-time text codecs is specified for 3G-324M and MTSI. For video there is a difference in levels, which mainly affects the maximum bitrate. Both video codec level and maximum bitrate can be specified as part of the call setup negotiation (see clause 12.2.5). Thus, it is very likely that the MTSI terminal and UE can agree on a common codec end-to-end without the need for MGW transcoding.

If a common codec is not found and the MGW does not support transcoding between any of the supported codecs, then the MGW may drop the unsupported media component. If the speech part cannot be supported, then the connection should not be set up.

12.2.2.2 Text

The CTM coding format defined in 3GPP TS 26.226 [52] is used for real time text in CS calls. In order to arrange inter-working, a transcoding function between CTM and RFC 4103 is required in the media gateway. A buffer shall be used for rate adaptation between receiving text from a real-time text transmitter according to the present document and transmitting to a CTM receiver. A gateway buffer of 2K characters is considered sufficient according to clause 13.2.4 in EG 202 320 [51].

Both CTM and RFC 4103 make use of ITU-T Recommendation T.140 presentation and character coding. Therefore inter-working is a matter of payload packetization and CTM modulation/demodulation.

A channel for real-time text is specified in ITU-T H.324. Also for this case, presentation and coding is specified according to ITU-T Recommendation T.140. Inter-working is a matter of establishing the text transport channels and moving the text contents between the two transport levels.

12.2.3 Payload format

See clause 7.4 of the present document.

12.2.4 Media gateway trans-packetization

12.2.4.1 General

The MGW shall offer conversion between H.223 as used in 3G-324M on the CS side and RTP as used in IMS. This clause contains a list inter-working functionalities that should be included.

12.2.4.2 Speech de-jitter buffer

The MGW should use a speech de-jitter buffer in the direction IMS to CS with sufficient performance to meet the 10 milliseconds maximum jitter requirement in clause 6.7.2 of ITU-T Recommendation H.324. H.324 specifies that transmission of each speech AL-SDU at the H.223 multiplex shall commence no later than 10 milliseconds after a whole multiple of the speech frame interval, measured from transmission of the first speech frame.

12.2.4.3 Video bitrate equalization

Temporary video rate variations can occur on the IMS side for example due to congestion. The video rate on the CS side, in contrast, is under full control of the CS side UE and the MGW.

During session setup, the MGW shall negotiate a video bitrate on the IMS side that allows all video bits to be conveyed to/from the CS link.

A buffer shall be maintained in the direction from the IMS to the CS side. The size of the buffer should be kept small enough to allow for a low end-to-end delay, yet large enough to conceal most network jitter on the IMS side. Temporary uneven traffic on the IMS side, beyond the handling capability of the buffer, should be handled as follows: if the buffer overflows, RTP packets should be dropped and the resulting loss and observed jitter should be reported by the means of an RTCP RR at the earliest possible sending time. The drop strategy may preferably be implemented media aware (i.e. favouring dropping predicted information over non-predicted information and similar techniques), or may be drop-head. If the buffer runs empty, the CS side should insert appropriate flag stuffing.

A buffer shall be maintained in the direction from the CS to the IMS side. The size of the buffer should be kept small enough to allow for a low end-to-end delay, but large enough to conceal most network jitter on the CS side. If the buffer overflows, then video bits must be dropped, preferably in a media-aware fashion, i.e. at GOB/slice/picture boundaries. MGCs may also take into account the type of media data, i.e. coded with or without prediction. If overflows occur frequently, the MGW may attempt to reduce the sending rate of the CS UE by employing H.245's FlowControlCommand. When the buffer runs empty, no activity is required on the IMS side.

If the bandwidth resources on the IMS side during a significant period of time drops below the limit where all video bits from the CS side can be forwarded, the MGW should drop the video component on the IMS side and change the CS call to a voice-only call [46]. The MGW should avoid dropping the entire call, so if the procedures in [46] are not available or feasible, the CS video call may be kept with the video component muted. If the video component was muted in the MGW for this reason and the available bandwidth on the IMS side increases, the MGW should restore the video component on the IMS side and un-mute the video on the CS side.

If the CS video call is changed to a voice-only call [46], the video component on the IMS side shall be dropped.

12.2.4.4 Data loss detection

If RTP packet loss is detected on input to the MGW at the IMS side, including losses caused by buffer-full condition as described above, corresponding H.223 AL-SDU sequence number increments should be made on the CS side to enable loss detection and proper concealment in the receiving CS UE.

If packet loss is detected on the CS side, e.g. through H.223 AL-SDU sequence numbers, those losses should be indicated towards the IMS side through corresponding RTP packet sequence number increments. The deliberate increments made for this reason will be visible in the RTCP RR from the MTSI UE and the MGW should take that into account when acting on RTCP RR from the MTSI UE, as the CS side losses are not related to the IMS network conditions.

12.2.4.5 Data integrity indication

This is mainly relevant in the direction from CS to IMS. The H.223 AL-SDUs include a CRC that forms an unreliable indication of data corruption. On the IMS side, no generic protocol mechanisms are available to convey this CRC and/or the result of a CRC check. The MGW shall discard any AL-SDUs which fail a CRC check and are not of a payload type that supports the indication of possible bit errors in the RTP payload header or data. If such payload type is in use, the MGW may forward corrupted packets, but in this case shall indicate the possible corruption by the means available in the payload header or data. One example is setting the Q bit of RFC 3267 [28] to 0 for AMR speech data that was carried in an H.223 AL-SDU with CRC indicating errors. Another example is setting the F bit of RFC 3984 [25] for H.264 NAL units that may contain bit errors.

The H.223 AL-SDU CRC is not fully fail-safe and it is therefore recommended that a MTSI terminal is designed to be robust and make concealment of corrupt media data, similar to the CS UE.

12.2.4.6 Packet size considerations

The same packet size and alignment requirements and considerations as defined in clause 7.5.2 of the present document and in 3GPP TS 26.111 [45] apply to the MGW, as it in that sense acts as a terminal and UE towards both the IMS and the CS side. Maximum available buffer size for packetization of media data may differ between IMS and CS UE and there currently exist no general means to signal this end-to-end. The maximumAl2SDUSize and maximumAl3SDUSize fields of the H223Capability member in H.245 TerminalCapabilitySet message have currently no counterpart in SIP/SDP. Thus, the MGW may have to segment data, especially video, in a non-favourable way. The number of such unfavourable segmentations should be kept to a minimum. Lacking general means for signalling, it is recommended to make use of available codec-specific packet-size signalling on the IMS side, such as the SDP receiver-capability parameter max-rcmd-nalu-size for H.264.

12.2.4.7 Setting RTP timestamps

In general, no explicit timestamps exist at the CS side. Even without transcoding functionality, the MGW may have to inspect and be able to interpret media data to set correct RTP timestamps.

12.2.4.8 Protocol termination

The MGW shall terminate the H.223 protocol at the CS side. Similarly, the MGW shall terminate RTP and RTCP at the IMS side.

12.2.4.9 Media synchronization

The MGW shall forward and translate the timing information between the IMS side (RTP timestamps, RTCP sender reports) and the CS side (H.245 message H223SkewIndication) to allow for media synchronization in the MTSI terminal and the CS UE. The MGW shall account for its own contribution to the skew in both directions. Note that transmission timing of H223SkewIndication and RTCP SR must be decoupled. H223SkewIndication has no timing restrictions, but is typically sent only once in the beginning of the session. RTCP SR timing is strictly regulated in RFC 3550 [9], RFC 4585[40], and clause 7.3. To decouple send timings, the time shift information conveyed in H223SkewIndication and RTCP SR must be kept as part of the MGW/MGCF session state. H223SkewIndication shall be sent at least once, and may be sent again when RTCP SR indicates a synchronization change. A synchronization change of less than 50 ms (value to be confirmed) should be considered insignificant and need not be signalled.

12.2.5 Session control

The MGCF shall offer translation between H.245 and SIP/SDP signalling according to 3GPP TR 29.863 [44] to allow for end-to-end capability negotiation.

12.3 GERAN/UTRAN CS inter-working

12.3.1 Codecs for media gateways

12.3.1.1 Speech

Media gateways offering speech communication between MTSI clients and non-MTSI clients operating in the CS domain in GERAN and UTRAN should support Tandem-Free Operation (TFO) according to 3GPP TS 28.062 [37], and Transcoder-Free Operation (TrFO), see 3GPP TS 23.153 [38].

MTSI media gateways offering speech communication and supporting TFO and/or TrFO shall support:

- AMR speech codec modes clauses 12.2, 7.4, 5.9 and 4.75 [11], [12], [13], [14] and source-controlled rate operation [15].
- Operation according to the UMTS_AMR_2 codec type with the Config-NB-Code=1 configuration as defined in [16].

MTSI media gateways should also support the other codec types and configurations as defined in [16].

When transmitting to the PS client, the media gateway shall be capable of restricting codec mode changes to be aligned to every other frame border, e.g. like UMTS_AMR_2 [16], and shall be capable of restricting codec mode changes to neighbouring codec modes within the negotiated codec mode set. The media gateway should be capable of changing codec mode aligned to every frame border and to any codec mode within the negotiated codec mode set. When receiving from the PS client, the media gateway shall allow codec mode changes at any frame border and to any codec mode within the negotiated codec mode set.

MTSI media gateways offering wideband speech communication at 16 kHz sampling frequency and supporting TFO and/or TrFO for wideband speech shall support:

- AMR wideband codec clauses 12.65, 8.85 and 6.60 [17], [18], [19], [20] and source controlled rate operation [21].

- Operation according to the UMTS_AMR_WB codec type with the Config-WB-code=0 configuration as defined in [16].

MTSI media gateways offering wideband speech communication at 16 kHz sampling frequency should also support the other codec types and configurations as defined in [16].

When transmitting to the PS client, the media gateway shall be capable of restricting codec mode changes to be aligned to every other frame border, e.g. like UMTS_AMR_WB [16], and shall be capable of restricting codec mode changes to neighbouring codec modes within the negotiated codec mode set. The media gateway should be capable of changing codec mode aligned to every frame border and to any codec mode within the negotiated codec mode set. When receiving from the PS client, the media gateway shall allow codec mode changes at any frame border and to any codec mode within the negotiated codec mode set.

MTSI terminals offering wideband speech communication shall also offer narrowband speech communications. When offering both wideband speech and narrowband speech communication, wideband shall be listed as the first payload type in the m line of the SDP offer (RFC 4566 [8]).

Requirements applicable to media gateways for DTMF events are described in Annex G.

12.3.2 RTP payload formats for media gateways

12.3.2.1 Speech

MTSI media gateways shall support the bandwidth-efficient payload format and should support the octet-aligned payload format. When offering both payload formats, the bandwidth-efficient payload format shall be listed before the octet-aligned payload format in the preference order defined in the SDP.

The MTSI media gateway shall use the parameters defined in table 12.1 during the session, unless the remote side does prevent it.

For all access technologies and for normal operating conditions, the MTSI media gateway should encapsulate the number of non-redundant speech frames in the RTP packets that corresponds to the ptime value defined in table 12.1. The MTSI media gateway may encapsulate more non-redundant speech frames in the RTP packet but shall not encapsulate more than 4 non-redundant speech frames in the RTP packets. The MTSI media gateway may encapsulate any number of redundant speech frames in an RTP packet but the length of an RTP packet, measured in ms, shall never exceed the maxptime value.

Table 12.1: Recommended encapsulation parameters

Access technology	Recommended encapsulation	ptime	maxptime when redundancy is not supported	maxptime when redundancy is supported
Unknown	1 non-redundant speech frame per RTP packet Max 4 or 12 speech frames in total depending on whether redundancy is supported	20	80	240
HSPA	1 non-redundant speech frame per RTP packet Max 4 or 12 speech frames in total depending on whether redundancy is supported	20	80	240
EGPRS	2 non-redundant speech frames per RTP packet Max 4 or 12 speech frames in total depending on whether redundancy is supported	40	80	240
GIP	1 to 4 non-redundant speech frames per RTP packet Max 12 speech frames in total	20, 40, 60 or 80	N/A	240

The SDP offer shall include an RTP payload type where octet-align=0 is defined or where octet-align is not specified and should include another RTP payload type with octet-align=1. MTSI media gateways offering wide-band speech shall offer these parameters and parameter settings also for the RTP payload types used for wide-band speech.

MTSI media gateways should support redundancy according to clause 9.

NOTE: Support of transmitting redundancy may be especially useful in the case an MTSI media gateway is aware of the used access technology and knows that the Generic Access technology is used.

12.4 PSTN

12.4.1 3G-324M

If 3G-324M is supported in the PSTN, then the inter-working can be made as specified in clause 12.2.

12.4.2 Text

PSTN text telephony inter-working with PS environments is described in ITU-T Recommendation H.248.2 [50]and further elaborated in EG 202 320 [51].

Text telephony modem tones are sensitive to packet loss, jitter and echo canceller behaviour. Therefore, conversion of modem based transmission of real-time text is best done at the border of the PSTN. If PSTN text telephone tones need to be carried audio coded in a PS network, considerations must be taken to carry them reliably as for example specified in ITU-T Recommendations V.151 [54] and V.152 [55].

When inter-working with PSTN text telephones, it must be considered that in PSTN most text telephone communication methods do not allow simultaneous voice and text transmission. An MTSI terminal indicating text capability shall not automatically initiate text connection efforts on the PSTN circuit. Instead, either a requirement for text support should be required from the MTSI terminal, active transmission of text from the MTSI terminal, or active transmission of text telephone tones from the PSTN terminal. See clause 13 of EG 202 320 [51].

Note that the primary goal of real-time text support in MTSI is not to offer a replica of PSTN text telephony functionality. On the contrary, real-time text in MTSI is aiming at being a generally useful mainstream feature, complementing the general usability of the Multimedia Telephony Service for IMS.

12.5 GIP inter-working

12.5.1 Text

RFC 4103 [53] and T.140 are specified as default real-time text codec in SIP telephony devices in RFC 4504 [53]. When GIP implements this codec, the media stream contents are identical for the two environments. Packetization will also in many cases be equal, while consideration must be taken to cope with different levels of redundancy and possible use of different media security and integrity measures.

12.6 TISPAN/NGN inter-working

12.6.1 Text

The codec and other considerations for real-time text described in the present document apply also to TISPAN/NGN. There are thus no inter-working considerations on the media level.

13 Supplementary services

13.1 General

In this section media layer behaviour is specified for relevant supplementary services. The supplementary services included in MTSI are described in 3GPP TS 24.173 [57]. The requirements on the codec support and the data transport are identical to those listed in clauses 5.2 and 7. These requirements are listed here due to the fact that there might be other media-influencing nodes in MTSI whose behaviour is not explicitly covered by other parts of the present document.

The recommended behaviour described in the following sections is valid for all session IP end-points; terminals, media gateways and other 3GPP network nodes acting as IP endpoints in MTSI sessions.

13.2 Media formats and transport

Any implementation of a supplementary service which affects media or media handling, e.g. such as media creation, media rendering and media manipulation, shall meet the same requirements as a terminal regarding codec support and codec usage. Hence, speech codecs shall be supported according to clause 5.2.1, video according to clause 5.2.2 and text according to clause 5.2.3.

Similarly, the configuration and the transport of the media in any implementation of a supplementary service which affects media or media handling shall be done according to clause 7.

13.3 Media handling in hold procedures

Whenever a supplementary service includes a hold procedure according to RFC 3264 [58], e.g. when using the HOLD supplementary service, the media flow is changed in terms of the session flow attribute (e.g. changing the session attribute "sendrecv" into "sendonly" or "recvonly" or "inactive" and then back again). When this occurs, any involved media-originating or media-terminating node should take measures to ensure that the transitions between the different media flow states in the session occur with minimal impact on the media quality.

When a full-duplex session has put the media flow on hold (see section 8.4 in RFC 3264 [58]), the media flow has been changed into a unidirectional flow through changing the session attribute into either "sendonly" or "recvonly". When resuming the session, it is restored to full duplex by changing the flow attributes back into "sendrecv" from "sendonly" and "recvonly". In this case, the encoder and decoder states in the clients may not be aligned and a state mismatch could occur. This would result in media quality degradation. Therefore, the following actions are recommended whenever the media session is not being put on hold anymore and the session is restored to full duplex:

- for speech media, the speech decoders should be reset;
- for video media, the video encoders should start the updated session with a full infra refresh even if the previously allocated encoders are still active and no infra refresh is scheduled to be sent.

Annex A (informative): Examples of SDP offers and answers

A.1 SDP offers for speech sessions initiated by terminal

This Annex includes several SDP examples for session setup for speech. SDP examples for sessions with speech and DTMF are shown in Annex G.

A.1.1 HSPA or unknown access technology

A.1.1.1 Only AMR-NB supported by MTSI terminal

In this example one RTP Payload Type (97) is defined for the bandwidth-efficient payload format and another RTP payload type (98) for the octet-aligned payload format. In this case, the terminal supports mode changes at any time, mode changes to any mode and mode change restrictions.

Table A.1.1: SDP example

SDP offer
m=audio 49152 RTP/AVP 97 98
a=rtpmap:97 AMR/8000/1
a=fmtp:97 mode-change-capability=2; max-red=220
a=rtpmap:98 AMR/8000/1
a=fmtp:98 mode-change-capability=2; max-red=220; octet-align=1
a=ptime:20
a=maxptime:240

Comments:

The UDP port number (49152) and the payload type numbers (97 and 98) are examples and the offerer is free to select other numbers within the restrictions of the UDP and RTP specifications. It is recommended to use the dynamic port numbers in the 49152 to 65535 range. RTP should use even numbers for RTP media and the next higher odd number for RTCP. It is however allowed to use any number within the registered port range 1 024 to 49 151. The receiver must be capable of using any combination of even and odd numbers for RTP and RTCP.

It is important that the terminal does not define any mode-set because then the answerer is free to respond with any mode-set that it can support. If the terminal would define mode-set to any value, then the answer only has the option to either accept it or reject it. The latter case might require several ping-pong between the end-points before they can reach an agreement on what mode set to use in the session. This would increase the setup time significantly. This is also one important reason for why the terminals must support the complete codec mode set of the AMR and AMR-WB codecs, because then a media gateway interfacing GERAN or UTRAN can immediately define the mode-set that it supports on the GERAN or UTRAN circuit switched access.

Since the terminal is required to support mode changes at any frame border and also to any mode in the received media stream, it does not set the mode-change-period and mode-change-neighbor parameters.

The mode-change-capability and max-red parameter are new in the updated AMR payload format [28]. With mode-change-capability=2, the terminal shows that it does support aligning mode changes every other frame and the answerer then knows that requesting mode-change-period=2 in the SDP answer will work properly. The max-red parameter indicates the maximum interval between a non-redundant frame and a redundant frame. Note that the maxptime and max-red parameters do not need to be synchronized.

The payload type for the bandwidth-efficient payload format (97) is listed before the payload type for the octet-aligned payload format (98) because it is the preferred one.

With the combination of ptime:20 and maxptime:240, the terminal shows that it desires to receive one speech frame per packet but can handle up to 12 speech frames per packet. Given the requirement that no more than 4 original speech frames can be encapsulated in one packet, the maxptime:240 setting means that redundancy with up to 8 redundant speech frames per packet is supported.

A.1.1.2 AMR and AMR-WB are supported by MTSI terminal

A.1.1.2.1 One-phase approach

The size of the SDP may become quite big, depending on how many configurations the terminal supports for different media. Therefore, the session setup may be divided into phases where the most desirable configurations are offered in the first phase. If the first phase fails, then the remaining configurations can be offered in a second phase.

In table A.1.2 an example is shown where a one-phase approach is used and where the SDP includes both AMR and AMR-WB and both the bandwidth-efficient and octet-aligned payload formats.

Table A.1.2: SDP example: one-phase approach

SDP offer
m=audio 49152 RTP/AVP 97 98 99 100
a=rtpmap:97 AMR-WB/16000/1
a=fmtp:97 mode-change-capability=2; max-red=220
a=rtpmap:98 AMR-WB/16000/1
a=fmtp:98 mode-change-capability=2; max-red=220; octet-align=1
a=rtpmap:99 AMR/8000/1
a=fmtp:99 mode-change-capability=2; max-red=220
a=rtpmap:100 AMR/8000/1
a=fmtp:100 mode-change-capability=2; max-red=220; octet-align=1
a=ptime:20
a=maxptime:240

Comments:

It is easy to imagine that the SDP offer can become quite large if the client supports many different configurations for one or several media.

A.1.1.2.2 Two-phase approach

Tables A.1.3 and A.1.4 show the same configurations as in table A.1.2 but when the SPD has been divided into 2 phases.

Table A.1.3: SDP example: 1st phase SDP offer

SDP offer
m=audio 49152 RTP/AVP 97 98
a=rtpmap:97 AMR-WB/16000/1
a=fmtp:97 mode-change-capability=2; max-red=220
a=rtpmap:98 AMR/8000/1
a=fmtp:98 mode-change-capability=2; max-red=220
a=ptime:20
a=maxptime:240

Table A.1.4: SDP example: 2nd phase SDP offer

SDP offer
m=audio 49152 RTP/AVP 97 98
a=rtpmap:97 AMR-WB/16000/1
a=fmtp:97 mode-change-capability=2; max-red=220; octet-align=1
a=rtpmap:98 AMR/8000/1
a=fmtp:98 mode-change-capability=2; max-red=220; octet-align=1
a=ptime:20
a=maxptime:240

Comments:

Many types of media and maybe even many different configurations for some or all media types, may give quite large SIP messages. When constructing the offer, the access type and the radio bearer(s) for the answerer are not yet known. To maintain a reasonable setup time, a 2-phase approach may be useful where the most desirable configurations are included in the 1st phase and the 2nd phase is entered only if all payload types for one media type are rejected.

There is however a drawback with the two-phase approach. If the 2nd phase is not entered, then a cell change that would require configurations from the 2nd phase SDP is likely to give long interruption times, several seconds, while the session parameters are re-negotiated.

A.1.2 EGPRS

In this example one RTP Payload Type (97) is defined for the bandwidth-efficient payload format and another RTP Payload Type (98) is defined for the octet-aligned payload format.

Table A.1.5: SDP example

SDP offer
m=audio 49152 RTP/AVP 97 98
a=rtpmap:97 AMR/8000/1
a=fmtp:97 mode-change-capability=2; max-red=220
a=rtpmap:98 AMR/8000/1
a=fmtp:98 mode-change-capability=2; max-red=220; octet-align=1
a=ptime:40
a=maxptime:240

Comments:

The only difference compared with the SDP offer for HSPA is ptime: 40. This definition is used to optimize capacity by reducing the amount of overhead that lower layers introduce. Defining ptime:20 will also work, but will be less optimal. Thus, when performing a cell change from HSPA to EGPRS, it is not an absolute necessity to update the session parameters immediately. It can be done after a while, which would also reduce the amount of SIP signalling if a UE is switching frequently between HSPA and EGPRS or some other access type.

It is recommended to set the max-red parameter to an even multiple of the ptime even though it is not required.

A.1.3 Generic Access

In this example one RTP Payload Type (97) is defined for the bandwidth-efficient payload format and another RTP Payload Type (98) is defined for the octet-aligned payload format.

Table A.1.6: SDP example

SDP offer
m=audio 49152 RTP/AVP 97 98
a=rtpmap:97 AMR/8000/1
a=fmtp:97 mode-change-capability=2; max-red=220
a=rtpmap:98 AMR/8000/1
a=fmtp:98 mode-change-capability=2; max-red=220; octet-align=1
a=ptime:80
a=maxptime:240

Comments:

In this case the terminal has detected that the load on the WLAN network is quite high and therefore ptime is set to 80. For other operating conditions, it could set ptime to 20, 40 or 60. This parameter may be updated during the session if the load of the WLAN network changes.

A.2 SDP offers for speech sessions initiated by media gateway

A.2.1 General

These examples show only SDP offers when the MTSI media gateway does not support the same configurations as for the MTSI terminal in clause A.1. A media gateway supporting the same configurations as for the examples in clause A.1 should create the same SDP offers.

A.2.2 MGW between GERAN UE and MTSI

This example shows the SDP offer when the call is initiated from GSM CS using the AMR with the {12.2, 7.4, 5.9 and 4.75} codec mode set. In this example, it is also assumed that only the bandwidth-efficient payload format is supported and that it will not send any redundant speech frames.

Table A.2.1: SDP example

SDP offer
m=audio 49152 RTP/AVP 97
a=rtpmap:97 AMR/8000/1
a=fmtp:97 mode-set=0,2,4,7; mode-change-period=2, \
mode-change-neighbor=1; mode-change-capability=2; max-red=0
a=ptime:20
a=maxptime:80

Comments:

Since the MGW only supports a subset of the AMR codec modes, it needs to indicate this in the SDP. The same applies for the mode change restrictions.

The broken a=fmtp line ("\") is in reality one single line in a real SDP.

A.2.3 MGW between legacy UTRAN UE and MTSI

This example shows the SDP offer when the call is initiated from legacy UTRAN CS mobile that only the AMR 12.2 mode. In this example, it is also assumed that only the bandwidth-efficient payload format is supported.

Table A.2.2: SDP example

SDP offer
m=audio 49152 RTP/AVP 97 a=rtpmap:97 AMR/8000/1 a=fmtp:97 mode-set=7; max-red=0 a=ptime:20 a=maxptime:20

Comments:

Since only one mode is supported, the mode-change-period, mode-change-neighbor and mode-change-capability parameters do not apply.

In this case it is advisable to not allow redundancy since the legacy UTRAN CS mobile does not support any lower rate codec modes and then redundancy would almost double the bitrate on the PS access side. Therefore, maxptime is set to 20 and max-red is set to 0.

If a mode-set with several codec modes was defined and if max-red and maxptime are set to larger values than what table A.1.8 shows, then redundancy is possible on the PS access side but not together with TFO.

A.2.4 MGW between CS UE and MTSI

This example shows the SDP offer when two mode sets are supported by the MGW.

Table A.2.3: SDP example

SDP offer

m=audio 49152 RTP/AVP 97 98
a=rtpmap:97 AMR/8000/1
a=fmtp:97 mode-set=0,2,4,7; mode-change-period=2, \\
  mode-change-neighbor=1; mode-change-capability=2; max-red=20
a=rtpmap:98 AMR/8000/1
a=fmtp:98 mode-set=0,3,5,6; mode-change-period=2, \\
  mode-change-neighbor=1; mode-change-capability=2; max-red=20
a=ptime:20
a=maxptime:80

Comments:

Redundancy up to 100 % is supported in this case since max-red is set to 20.

A.3 SDP answers to SDP speech session offers

A.3.1 General

This clause gives a few examples of possible SDP answers. The likelihood of these SDP answers may vary from case to case. It is impossible to cover all the possible variants and hence these examples were selected because they span the range quite well.

The SDP offers are included to clarify what is being answered.

A.3.2 SDP answer from an MTSI UE

These SDP offers and answers are likely when both UEs support AMR and AMR-WB and also both the bandwidth-efficient and the octet-aligned payload formats.

Table A.3.1: SDP example

SDP offer
m=audio 49152 RTP/AVP 97 98 99 100 a=rtpmap:97 AMR-WB/16000/1 a=fmtp:97 mode-change-capability=2; max-red=220 a=rtpmap:98 AMR-WB/16000/1 a=fmtp:98 mode-change-capability=2; max-red=220; octet-align=1 a=rtpmap:99 AMR/8000/1 a=fmtp:99 mode-change-capability=2; max-red=220 a=rtpmap:100 AMR/8000/1 a=fmtp:100 mode-change-capability=2; max-red=220; octet-align=1 a=ptime:20 a=maxptime:240
SDP answer
m=audio 49152 RTP/AVP 97 98 99 100 a=rtpmap:97 AMR-WB/16000/1 a=fmtp:97 mode-change-capability=2; max-red=220 a=rtpmap:98 AMR-WB/16000/1 a=fmtp:98 mode-change-capability=2; max-red=220; octet-align=1 a=rtpmap:99 AMR/8000/1 a=fmtp:99 mode-change-capability=2; max-red=220 a=rtpmap:100 AMR/8000/1 a=fmtp:100 mode-change-capability=2; max-red=220; octet-align=1 a=ptime:20 a=maxptime:240

Comments:

Since both UEs support the same configurations, it is likely that the answer is identical to the offer. The conclusion from this offer-answer process is that AMR-WB will be used during the session with RTP Payload Type 97.

Even though both UEs support all codec modes, it is desirable to mainly use the codec modes from the AMR {12.2, 7.4, 5.9 and 4.75} and AMR-WB {12.65, 8.85 and 6.60} mode sets because the transport layer functions are optimized for these modes.

For similar reasons it is also desirable to encapsulate only 1 speech frame per packet, even though both UEs support receiving several frames per packet.

A.3.3 SDP answer from an MTSI UE supporting only AMR

These SDP offers and answers are likely when the answering UE support only AMR.

Table A.3.2: SDP example

SDP offer
m=audio 49152 RTP/AVP 97 98 99 100 a=rtpmap:97 AMR-WB/16000/1 a=fmtp:97 mode-change-capability=2; max-red=220 a=rtpmap:98 AMR-WB/16000/1 a=fmtp:98 mode-change-capability=2; max-red=220; octet-align=1 a=rtpmap:99 AMR/8000/1 a=fmtp:99 mode-change-capability=2; max-red=220 a=rtpmap:100 AMR/8000/1 a=fmtp:100 mode-change-capability=2; max-red=220; octet-align=1 a=ptime:20 a=maxptime:240
SDP answer
m=audio 49152 RTP/AVP 99 100 a=rtpmap:99 AMR/8000/1 a=fmtp:99 mode-change-capability=2; max-red=220 a=rtpmap:100 AMR/8000/1 a=fmtp:100 mode-change-capability=2; max-red=220; octet-align=1 a=ptime:20 a=maxptime:240

Comments:

In the answer, RTP Payload Types 97 and 98 have been removed since AMR-WB is not supported.

A.3.4 SDP answer from an MTSI UE camping on EGPRS

In this case the answering UE is using EGPRS access.

Table A.3.3: SDP example

SDP offer
m=audio 49152 RTP/AVP 97 98 99 100 a=rtpmap:97 AMR-WB/16000/1 a=fmtp:97 mode-change-capability=2; max-red=220 a=rtpmap:98 AMR-WB/16000/1 a=fmtp:98 mode-change-capability=2; max-red=220; octet-align=1 a=rtpmap:99 AMR/8000/1 a=fmtp:99 mode-change-capability=2; max-red=220 a=rtpmap:100 AMR/8000/1 a=fmtp:100 mode-change-capability=2; max-red=220; octet-align=1 a=ptime:20 a=maxptime:240
SDP answer
m=audio 49152 RTP/AVP 97 98 99 100 a=rtpmap:97 AMR-WB/16000/1 a=fmtp:97 mode-change-capability=2; max-red=200 a=rtpmap:98 AMR-WB/16000/1 a=fmtp:98 mode-change-capability=2; max-red=200; octet-align=1 a=rtpmap:99 AMR/8000/1 a=fmtp:99 mode-change-capability=2; max-red=200 a=rtpmap:100 AMR/8000/1 a=fmtp:100 mode-change-capability=2; max-red=200; octet-align=1 a=ptime:40 a=maxptime:240

Comments:

The answering UE responds that it desires to receive 2 frames encapsulated in each packet. It will however send with 1 frame per packet since the offering UE desires to receive this format. A future SIP UPDATE may change this so that 2 frames per packet are used in both directions.

The answering UE also responds with max-red defined to 200 ms since this is the closes multiple of the desired frame aggregation. It should however be noted that it is not a requirement to define max-red to be a multiple of ptime, but it is recommended to do so.

A.3.5 SDP answer from MGW supporting only one codec mode set for AMR and AMR-WB each

In this case the MGW supports only one codec mode set for AMR, {12.2, 7.4, 5.9 and 4.75}, and one codec mode set for AMR-WB, {12.65, 8.85 and 6.60}. The MGW also only supports the bandwidth-efficient payload format.

Table A.3.4: SDP example

SDP offer (from UE on HSPA)
m=audio 49152 RTP/AVP 97 98 99 100 a=rtpmap:97 AMR-WB/16000/1 a=fmtp:97 mode-change-capability=2; max-red=220 a=rtpmap:98 AMR-WB/16000/1 a=fmtp:98 mode-change-capability=2; max-red=220; octet-align=1 a=rtpmap:99 AMR/8000/1 a=fmtp:99 mode-change-capability=2; max-red=220 a=rtpmap:100 AMR/8000/1 a=fmtp:100 mode-change-capability=2; max-red=220; octet-align=1 a=ptime:20 a=maxptime:240
SDP answer (from MGW)

m=audio 49152 RTP/AVP 97 99
a=rtpmap:97 AMR-WB/16000/1
a=fmtp:97 mode-set=0,1,2; mode-change-period=2, mode-change-neighbor=1; \\
  mode-change-capability=2; max-red=0
a=rtpmap:99 AMR/8000/1
a=fmtp:99 mode-set=0,2,4,7; mode-change-period=2, mode-change-neighbor=1; \\
  mode-change-capability=2; max-red=0
a=ptime:20
a=maxptime:80

Comments:

The MGW is allowed to define the mode-set parameter since the UE did not define it. Thereby, it is possible to avoid several SDP offers and answers.

Since the UE has defined that it does support restrictions in mode changes, the MGW can safely set the mode-change-period and mode-change-neighbor parameters.

In this example, the MGW also does not support redundancy so it sets max-red to zero.

A.3.6 SDP answer from UE on HSPA for session initiated from MGW interfacing UE on GERAN

This example shows the offers and answers for a session between a GERAN CS client, through a media gateway, and a MTSI client.

Table A.3.5: SDP example

SDP offer (from MGW)
m=audio 49152 RTP/AVP 97
a=rtpmap:97 AMR/8000/1
a=fmtp:97 mode-set=0,2,4,7; mode-change-period=2, \
mode-change-neighbor=1; mode-change-capability=2; max-red=0
a=ptime:20
a=maxptime:20
SDP answer (from UE)
m=audio 49152 RTP/AVP 97
a=rtpmap:97 AMR/8000/1
a=fmtp:97 mode-set=0,2,4,7; mode-change-period=2, \
mode-change-neighbor=1; mode-change-capability=2; max-red=0
a=ptime:20
a=maxptime:240

Comments:

The media gateway offers only a restricted mode set since it cannot support anything else. The MTSI client has to accept this, if it wants to continue with the session setup.

This example also shows that the media gateway want to receive 1 frame per packet. The maxptime parameter is therefore set to 20. With max-red set to 0 the media gateway also shows that it will not send redundancy. The MTSI client can support receiving up to 12 frames per packet. It therefore set the maxptime parameter to 240.

The UE detects that the media gateway does not want to receive redundancy and therefore sets max-red to 0.

A.4 SDP offers and answers for video sessions

A.4.1 H.263 and MPEG-4 Visual

In the following example the SDP offer includes two video codec options:

Table A.4.1: Example SDP offer for H.263 and MPEG-4 Part 2 video

SDP offer
m=video 49154 RTP/AVP 99 100
b=AS:92
a=rtpmap:99 H263-2000/90000
a=fmtp:99 profile=0;level=45
a=rtpmap:100 MP4V-ES/90000
a=fmtp:100 profile-level-id=9; \
config=000001b009000001b509000001000000012000845d4c282c2090a28f

The two options in table 4.1 are associated with the RTP Payload Type numbers 99 and 100. The first offer includes ITU-T Recommendation H.263 Profile 0 (Baseline) at level 45, which supports bitrates up to 128 kbps and maximum QCIF picture formats at 15 Hz. The second offer is MPEG-4 Visual (Part 2) Simple profile at level L0b, which also supports bitrates up to 128 kbps and QCIF at 15 Hz. Here profile-level-id=9 represents Simple profile at level L0b and may be used for negotiation, whereas the config parameter gives the configuration of the MPEG-4 Visual bit stream and is not used for negotiation. The bandwidth (including IP, UDP and RTP overhead) for video is 92 kbps.

The broken a=fmtp line ("\") is in reality one single line in a real SDP.

An example SDP answer to the offer is given below.

Table A.4.2: Example SDP answer

SDP answer
m=video 49154 RTP/AVP 99
b=AS:48
a=rtpmap:99 H263-2000/90000
a=fmtp:99 profile=0;level=10

The answer includes only the H.263 codec. The responding client has restricted the video bandwidth to 48 kbps and restricted the H.263 level to 10 which supports bitrates up to 64 kbps. The offerer should not have a problem with a reduced bitrate as support for level 45 implies the support of level 10 as well.

A.4.2 H.264/AVC with H.263 as fallback

In this example the SDP offer includes H.264/AVC with H.263 as fallback.

Table A.4.3: Example SDP offer for H.264/AVC with H.263 as fallback

SDP offer
m=video 49154 RTP/AVP 99 100
b=AS:48
a=rtpmap:99 H264/90000
a=fmtp:99 packetization-mode=0;profile-level-id=42e00a; \
sprop-parameter-sets=J0LgCpWgsToB/UA=,KM4Gag==
a=rtpmap:100 H263-2000/90000
a=fmtp:100 profile=0;level=10

The first (preferred) offer is H.264/AVC. The packetization-mode parameter indicates single NAL unit mode. This is the default mode and it is therefore not necessary to include this parameter (see RFC 3984). The profile-level-id parameter indicates Baseline profile at level 1, which supports bitrates up to 64 kbps. It also indicates, by using so-called constraint-set flags, that the bit stream can be decoded by any Baseline, Main or Extended profile decoder. The third parameter, sprop-parameter-sets, includes base-64 encoded sequence and picture parameter set NAL units that are referred by the video bit stream. The sequence parameter set used here includes syntax that specifies the number of re-ordered frames to be zero so that latency can be minimized. The second offer in the SDP is H.263 Profile 0 (Baseline) at level 10. It is used here as a fallback in case the other client does not support H.264/AVC. The bandwidth (including IP, UDP and RTP overhead) for video is restricted to 48 kbps.

An example SDP answer to the offer is given below.

Table A.4.4: Example SDP answer

SDP answer
m=video 49154 RTP/AVP 99 b=AS:48 a=rtpmap:99 H264/90000 a=fmtp:99 packetization-mode=0;profile-level-id=42e00a; \ sprop-parameter-sets=J0LgCpWgsToB/UA=,KM4Gag==

The responding client is capable of using H.264/AVC and has therefore removed the fallback offer H.263. As the offer already indicated the lowest level (level 1) of H.264/AVC as well as the minimum constraint set, there is no room for further negotiation of profiles and levels. However, the bandwidth could be constrained further by reducing the bandwidth in b=AS.

A.5 SDP offers for text

A.5.1 T.140 with and without redundancy

An offer to use T.140 real-time text may be realized by using SDP according to the following example in session setup or for addition of real-time text during a session.

Table A.5.1: Example SDP offer for T.140 real-time text

SDP offer
m=text 53490 RTP/AVP 100 98 a=rtpmap:100 red/1000/1 a=rtpmap:98 t140/1000/1 a=fmtp:100 98/98/98

The example in table A.5.1 shows that RTP payload type 98 is used for sending text without redundancy, whereas RTP payload type 100 is used for sending text with 200 % redundancy.

A.6 SDP example with bandwidth information

This clause gives an example where the bandwidth modifiers have been included in the SDP offer.

Table A.6.1: SDP example with bandwidth information

SDP offer
v=0 o=Example_SERVER 3413526809 0 IN IP4 server.example.com s=Example of AS, TIAS and maxprate in MTSI c=IN IP4 aaa.bbb.ccc.ddd b=AS:78 m=audio 49152 RTP/AVPF 97 98 b=AS:30 a=rtpmap:97 AMR/8000/1 a=fmtp:97 mode-change-capability=2; max-red=160 a=rtpmap:98 AMR/8000/1 a=fmtp:98 mode-change-capability=2; max-red=160; octet-align=1 a=ptime:20 a=maxptime:240 m=video 49154 RTP/AVPF 99 b=AS:48 a=rtpmap:99 MP4V-ES/90000 a=fmtp:99 profile-level-id=8; \\   config=000001B008000001B50900000101000001200088400668 2C2090A21F

The broken a=fmtp line ("\") is in reality one single line in a real SDP.

A.7 SDP examples with "3gpp_sync_info" attribute

A.7.1 Synchronized streams

In the example given below in table A.7.1, streams identified with "mid" attribute 1 and 2 are to be synchronized (default operation if the "3gpp_sync_info" attribute is absent).

Table A.7.1: SDP example with requirement on synchronization

SDP offer
v=0 o=Laura 289083124 289083124 IN IP4 one.example.com t=0 0 c=IN IP4 224.2.17.12/127 a=group:LS 1 2 a=3gpp_sync_info:Sync m=audio 30000 RTP/AVP 0 a=mid:1 m=video 30002 RTP/AVP 31 a=mid:2 m=audio 30004 RTP/AVP 2 i=This media stream contains the Spanish translation a=mid:3

A.7.2 Nonsynchronized streams

The SDP in table A.7.2 gives an example of the usage of "3gpp_sync_info" attribute at media level. In this example, the MPEG-4 video stream should not be synchronized with any other media stream in the session.

Table A.7.2: SDP example with no requirement on synchronization

SDP offer
v=0 o=Laura 289084412 2890841235 IN IP4 123.124.125.1 s=Demo c=IN IP4 123.124.125.1 m=video 6000 RTP/AVP 98 a=rtpmap:98 MP4V-ES/90000 a=3gpp_sync_jitter:No Sync m=video 5000 RTP/AVP 99 a=rtpmap 99 H263-2000/90000 m=audio 7000 RTP/AVP 100 a=rtpmap:100 AMR

A.8 SDP example with QoS negotiation

This clause gives an example of an SDP interchange with negotiated QoS parameters.

Table A.8.1: SDP example with QoS negotiation

SDP offer from UE A to B
v=0 o=Example_SERVER 3413526809 0 IN IP4 server.example.com s=Example of using AS to indicate negotiated QoS in MTSI c=IN IP4 aaa.bbb.ccc.ddd b=AS:78 m=audio 49152 RTP/AVPF 97 98 b=AS:30 a=rtpmap:97 AMR/8000/1 a=fmtp:97 mode-change-capability=2; max-red=160 a=rtpmap:98 AMR/8000/1 a=fmtp:98 mode-change-capability=2; max-red=160; octet-align=1 a=ptime:20 a=maxptime:240 m=video 49154 RTP/AVPF 99 b=AS:48 a=rtpmap:99 MP4V-ES/90000 a=fmtp:99 profile-level-id=8; \\ config=000001B008000001B5090000010100000120008840066 82C2090A21F

SDP from UE B to A in SIP UPDATE message
v=0 o=Example_SERVER2 34135268010 IN IP4 server2.example.com s=Example of using AS to indicate negotiated QoS in MTSI c=IN IP4 aaa.bbb.ccc.ddd b=AS:60 m=audio 49252 RTP/AVPF 97 98 b=AS:30 a=rtpmap:97 AMR/8000/1 a=fmtp:97 mode-change-capability=2; max-red=160 a=rtpmap:98 AMR/8000/1 a=fmtp:98 mode-change-capability=2; max-red=160; octet-align=1 a=ptime:20 a=maxptime:240 m=video 49254 RTP/AVPF 99 b=AS:30 a=rtpmap:99 MP4V-ES/90000 a=fmtp:99 profile-level-id=8; \ config=000001B008000001B50900000101000001200088400668 2C2090A21F
SDP from UE A to B in 200/OK RESPONSE to UPDATE message
v=0 o=Example_SERVER 3413526809 0 IN IP4 server.example.com s=Example of using AS to indicate negotiated QoS in MTSI c=IN IP4 aaa.bbb.ccc.ddd b=AS:78 m=audio 49152 RTP/AVPF 97 98 b=AS:30 a=rtpmap:97 AMR/8000/1 a=fmtp:97 mode-change-capability=2; max-red=160 a=rtpmap:98 AMR/8000/1 a=fmtp:98 mode-change-capability=2; max-red=160; octet-align=1 a=ptime:20 a=maxptime:240 m=video 49154 RTP/AVPF 99 b=AS:48 a=rtpmap:99 MP4V-ES/90000 a=fmtp:99 profile-level-id=8; \ config=000001B008000001B50900000101000001200088400668 2C2090A21F

The example in table A.8.1 shows an SDP exchange that reflects the signalling of negotiated QoS during initial session setup when there is only one PDP context for the whole session. When UE B gets a different negotiated QoS than what was indicated in the offer from A, it sends an UPDATE message to UE A indicating the negotiated QoS, UE A responds with its negotiated QoS value to B.

The broken a=fmtp line ("\") is in reality one single line in a real SDP.

A.9 SDP offer/answer regarding the use of non-compound RTCP

This example shows the offers and answers for a session between two clients controlling the use of non-compound RTCP.

Table A.9.1: SDP example for non-compound RTCP

SDP offer
m=audio 49152 RTP/AVPF 97 98 a=rtcp-fb:* trr-int 5000; ncp a=rtpmap:97 AMR/8000/1 a=fmtp:97 mode-change-capability=2; max-red=220 a=rtpmap:98 AMR/8000/1 a=fmtp:98 mode-change-capability=2; max-red=220; octet-align=1 a=ptime:20 a=maxptime:240

Comments:

This example allows the use of non-compound RTCP (attribute ncp) for the adaptation feedback. Moreover the minimum interval between two regular compound RTCP packets is set to 5000 milliseconds.

Annex B (informative): Examples of adaptation scenarios

B.1 Video bitrate adaptation

It is recommended in clauses 7.3.3 and 10.3 that a video sender adapts its video output rate based on RTCP reports and TMMBR messages. The following example illustrates the usage:

EXAMPLE:

1. A video session is established at 100kbps. 5kbps is allocated for RTCP and trr-int is set to 500 ms. This allows an end-point to send regular RTCP reports with an average 500 ms interval consuming less than 5 kbps for RTCP. At the same time it allows the end-point to send an early RTCP event packet and then send the next one already after 800 ms instead of after 1 000 ms.

2. The receiver is now subject to a reduced bandwidth, e.g. 60 kbps, due to handover to a different cell. The network indicates the reduced bandwidth to the receiver. The receiver generates a TMMBR message to inform the sender of the new maximum bitrate, 60 kbps.

3. The sender receives the TMMBR message, adjusts its output bitrate and sends a TMMBN message back.

4. The receiver sends a SIP UPDATE message to the sender indicating 60 kbps

5. The receiver travels into an area with full radio coverage. A new bandwidth of 100 kbps is negotiated with the network. It sends a SIP UPDATE message for 100 kbps.

6. The sender receives the SIP UPDATE message, and adjusts its output bitrate.

Annex C (informative): Example adaptation mechanism for speech

C.1 Example of feedback and adaptation for speech

C.1.1 Introduction

This annex gives the outline of possible example adaptation implementations that make use of adaptation signalling for speech as described in section 10.2. Several different adaptation implementations are possible and the examples shown in this section are not to be seen as a set of different adaptive schemes excluding other designs. Implementers are free to use these examples or to use any other adaptation algorithms. The examples are only based on measured packet losses whereas a real implementation is free to use other adaptation triggers. The purpose of the section is to show a few different examples of how receiver state machines can be used both to control the signalling but also to control the signalling requests. Notice that the endpoints can have different implementations of the adaptation state machines.

The annex is divided into three sections:

- Signalling considerations - Implementation considerations on the signalling mechanism; the signalling state machine.
- Adaptation state machines - Three different examples of adaptation state machines either using the full set of adaptation dimensions or a subset thereof.
- Other issues and solutions - Default actions and lower layer triggers.

In this annex, a *media receiver* is the receiving end of the media flow, hence the *request sender* of any adaptation request. A *media sender* is the sending entity of the media, hence the *request receiver* of the adaptation request. The three different adaptation mechanisms available; bit-rate, packet-rate and error resilience, represents different ways to adapt to current transport characteristics:

- Bit-rate adaptation. Reducing the bit-rate is in all examples shown in this section the first action done whenever a measurement indicating that action is needed to further optimize the session quality. A bit-rate reduction will reduce the utilization of the network resources to transmit the data. In the radio case, this would reduce the required transmission power and free resources either for more data or added channel coding. It is reasonable to assume, also consistent with a proper behaviour on IP networks, that a reduction of bit-rate is a valid first measure to take whenever the transport characteristics indicate that the current settings of the session do not provide an optimized session quality.
- Packet-rate adaptation. In some of the examples, packet-rate adaptation is a second measure available to further adapt to the transport characteristics. A reduction of packet rate will in some cases improve the session quality, e.g. in transmission channels including WLAN. Further, a reduction of packet rate will also reduce the protocol overhead since more data is encapsulated into each RTP packet. Although robust header compression (RoHC) can reduce the protocol overhead over the wireless link, the core network will still see the full header and for speech data, it consists of a considerable part of the data transmitted. Hence, packet-rate adaptation serves as a second step in reducing the total bit-rate needed for the session.
- Error resilience. The last adaptive measure in these examples is the use of error resilience measures, or explicitly, application level redundancy. Application level redundancy does not reduce the amount of bits needed to be transmitted but instead transmit the data in a more robust way. Application level redundancy should only be seen as a last measure when no other adaptation action has succeeded in optimizing the session quality sufficiently well. For most normal use cases, application level redundancy is not foreseen to be used, rather it serves as the last resort when the session quality is severely jeopardized.

C.1.2 Signalling state considerations

The control of the adaptation signalling can by itself be characterized as a state machine. The implementation of the state machine is in the decoder and each endpoint has its own implementation. The decoder sends requests as described in clause 10.2 to the encoder in the other end.

The requests that are transmitted can be queued up in a send buffer to be transmitted the next time an RTCP-APP packet is to be sent. Hence, a sender might receive one, two or all three receiver requests at the same time. It should not expect any specific order of the requests. A receiver shall not send multiple requests of the same type in the same RTCP-APP packet. Transmission of the requests should preferably be done immediately using the AVPF early mode but in some cases it may be justified to delay the transmission a limited time or until the next DTX period in order to minimize disturbance on the RTP stream, in the latter case monitoring of the RTP stream described below must take the additional delay into account.

To summarize:

- A request can be sent immediately (alone in one RTCP-APP packet) but the subsequent RTCP-APP packet must follow the transmission rules for RTCP.
- RTCP-APP packets may be delayed until the next DTX period.

Reception of the transmitted RTCP-APP packets is not guaranteed. Similar to the RTP packets, the RTCP packets might be lost due to link losses. Monitoring that the adaptation requests are followed can to be done by means of inspection of the received RTP stream.

For various reasons the requests might not be followed even though they received successfully by the other end. This behaviour can be seen in the following ways:

- Request completely ignored: An example is a request for 1 frame/packet which might be rejected as the endpoint decides that the default mode of operation 2 frames/packet or more and a frame aggregation reduction compared to the default state is not allowed.
- Request partially followed: An example here is when no redundancy is received and a request for 100 % redundancy with 1 extra frame offset is made which may be realized by the media sender as 100 % redundancy with no extra offset. Another example is when a request for 5.9 kbps codec rate is sent and it is realized as e.g. 6.7 kbps codec rate. Table C.1 displays how the requests and realizations are grouped. E.g. it can be seen (if Ninit =1) that a request for 3 frames per packets realized as 2 frames per packet is considered to be fulfilled.

Table C.1: Distinction of different settings for frame aggregation, redundancy and codec mode settings

Codec rate	Frame aggregation	Redundancy
Highest rate in mode set	N_{init} frame per packet	No redundancy
All other codec rates	≥ N_{init}+1 frames per packet	≥ 100 % redundancy , arbitrary offset

In table C.1 above N_{init} is 1 in most cases which corresponds to 1 frame per packet. In certain cases N_{init} might have another value, one such example is E-GPRS access where N_{init} may be 2. N_{init} is given by the ptime SDP attribute.

Note that special care in the monitoring should be taken when DTX is used as DTX SID update packets are normally not aggregated or transmitted redundant. Important is also that it takes at least one roundtrip before the effect of a request is seen in the RTP flow, if transmission of RTCP is delayed due to e.g. bandwidth requirements this extra delay must also be taken into account in the monitoring.

If the requests are not followed as requested, the request should not be repeated infinitely as it will increase the total bit-rate without clear benefit. In order to avoid such behaviour the following recommendations apply:

- Partially fulfilled requests should be considered as obeyed.
- If a new request is not fulfilled within T_RESPONSE ms, the request is repeated again with a delay between trials of 2*T_RESPONSE ms. If the three attempts have been made without sender action, it should be assumed that the request cannot be fulfilled. In this case, the adaptation state machine will stay in the previous state or in a state that matches the current properties (codec mode, redundancy, frame aggregation). Any potential mismatch between define states in the adaptation state machine and the current properties of the media stream should resolved by the request sender.
- The default mode of operation for a client if the RTCP bandwidth for the session is greater than zero is that the requests received should be followed. Ignoring requests should be avoided as much as possible. However, it is required that any signalling requests are aligned with the agreed session parameters in the SDP.

In some cases the adaptation state machine may go out-of-synch with the received RTP stream. Such cases may occur if e.g. the other endpoint makes a reset. These special cases can be sensed, e.g. through a detection of a large gap in timestamp and/or sequence number. The state machine should then reset to the default state and start over again.

The signalling state machine has three states according to table C.2.

Table C.2: Signalling state machine states

State	Description
T1	Idle state: This is the default state of the signalling state machine. The signalling state should always return here after a state transition and when it has been detected that the media sender has followed the request, either completely or partially. The signalling state machine remains in this state as long as the selected adaptation is "stable", i.e. as long as the adaptation measures are appropriate for the current operating conditions. When it has been detected that the operating conditions has changed so much that the current adaptation measures are no longer appropriate then the adaptation function triggers a request signalling and the signalling state machine goes to state T2.
T2	In this state, the received RTP stream is monitored to verify that the properties of a given adaptation state (redundancy, frame aggregation and codec mode) are detected in the received RTP stream. If necessary, some of the requests are repeated maximum 3 times. If any of the properties is considered to be not fulfilled, the signalling state machine enters state T3.
T3	In this state, the properties of the RTP stream (redundancy, frame aggregation and codec rate) is reverted back to the properties of the last successful state and a new state transition is tested in T2, or alternatively the adaptation state is set to the state that matches the current properties (codec mode, redundancy, frame aggregation).

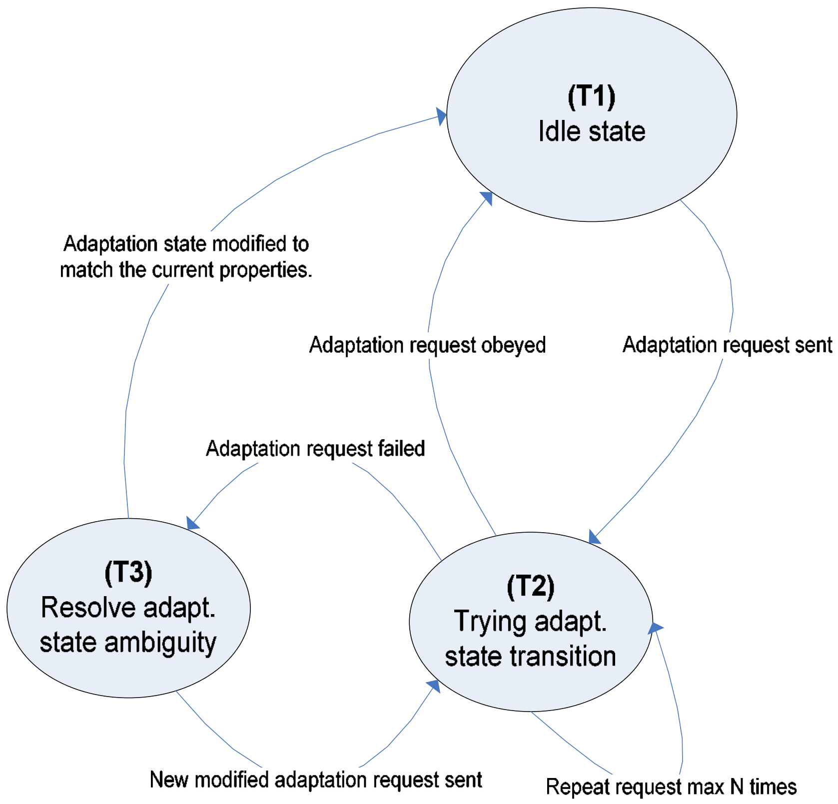


Figure C.1: Signalling state machine, implemented in order to ensure safe adaptation state transitions

C.1.3 Adaptation state machine implementations

C.1.3.1 General

The example adaptation state machines shown in this section are different realizations of the control algorithm for the adaptation requests. Note that this does not include how the actual signalling should be done but how various triggers will result in the transmission of different requests.

The example adaptation state machines make use of the signalling state machine outlined in clause B.2. Common to all adaptation state machines is that it is possible to implement all versions in the same code and just exclude appropriate states depending on desired mode of operation. All examples can transit between a number of states (denoted S1…S4). In these examples, it is assumed that the codec is AMR-NB and that it uses two coding rates (AMR 12.2 and AMR 5.9). However, this is not a limitation of the adaptation mechanism by itself. It is only the scenario used in these examples.

Since the purpose of the adaptation mechanism is to improve the quality of the session, any adaptation signalling is based upon some trigger; either a received indication or a measurement. In the case of a measurement trigger, it is important to gather reliable statistics. This requires a measurement period which is sufficiently long to give a reliable estimation of the channel quality but also sufficiently short to enable fast adaptation. For typical MTSI scenarios on 3GPP accesses, a measurement period in the order of 100 packets is recommended. Further, in order to have an adaptation control which is reliable and stable, a hangover period is needed after a new state has been entered (typically 100 to 200 packets). An even longer hangover period is suitable when transiting from an error resilient state or a reduced rate into the default, normal state. In the below examples, it is assumed that the metric used in the adaptation is the packet loss rate measured on the application layer. It is possible to use other metrics such as lower layer channel quality metrics.

Note that mode change requests must follow the rules outlined in clause 5.2.1.

The example solution is designed based on the following assumptions:

- When the packet loss rate increases, the adaptation should:
    - First try with a lower codec mode rate, i.e. bit-rate back off.
    - If this does not improve the situation, then one should try with packet rate back-off by increasing the frame aggregation.
    - If none of these methods help, then application layer redundancy should be added to save the session.
- When the packet loss rate increases, one should try to increase the bit rate in a "safe" manner. This is done by probing for higher bit rates by adding redundancy.
- The downwards adaptation, towards lower rates and redundancy, should be fast while the upwards adaptation should be slow.
- Hysteresis should be used to avoid oscillating behaviour between two states.

A description of the different states and what trigger the transition into the respective state is given in table C.3.

Table C.3: Adaptation state machine states and their meaning

State	Description
S1	Default/normal state: Good channel conditions. This state has the properties: • Codec rate: Highest mode in mode set. • Frame aggregation: Equal to the ptime value in the agreed session parameters. • Redundancy: 0%.
S2	In this state the encoding bit-rate and the packet rate is reduced. The state is divided into 2 sub states (S2a and S2b). In state S2a the codec rate is reduced and in state S2b the packet rate is also reduced (the frame aggregation is increased). State S2a may also involve a gradual decrease of the codec-rate in order to be in agreement with the session parameters. If no restrictions are in place regarding mode changes (i.e. such as only allowing changing to a neighbouring mode), it changes bit-rate to the target reduced bit-rate directly. If restrictions are in place, several mode changes might be needed. This state has the properties: • Codec rate: Any codec rate except the highest rate in mode set, preferably a codec rate that is roughly half the highest rate. • Frame aggregation: o S2a: Equal to the ptime value in the agreed session parameters. o S2b: ptime+N*20ms where N > 1, limited by max-ptime. • Redundancy: 0%.
S3	This is an interim state where the total bit-rate and packet rate is roughly equal to state S1. 100% redundancy is used with a lower codec mode than S1. This is done to probe the channel band-width with a higher tolerance to packet loss to determine if it is possible to revert back to S1 without significantly increase the packet loss rate. This state has the properties: • Codec rate: Any codec rate except the highest rate in mode set, preferably a codec rate that is roughly half the highest rate, target total rate (with redundancy) should be roughly the same as in S1. • Frame aggregation: Equal to the ptime value in the agreed session parameters. • Redundancy: 100%.
S4	In this state the encoding bit-rate is reduced (the same bit-rate as in S2) and redundancy is turned on. Optionally also the packet rate is kept the same as in state S2. This state has the properties: • Codec rate: Any codec rate except the highest rate in mode set, preferably a codec rate that is roughly half the highest rate. • Frame aggregation: Equal to the ptime value in the agreed session parameters. • Redundancy: 100%, possibly with offset.

The parameters and other definitions controlling the behaviour of the adaptation state machine are described in table C.4. Example values are also shown, values which give good performance on a wide range of different channel conditions.

Table C.4: State transition definitions, thresholds and temporal adaptation control parameters

Parameter	Value/meaning	Comment
PLR_1	3 %	
PLR_2	1 %	
PLR_3	2 %	
PLR_4	10 %	
N_INHIBIT	1 000 frames	A random value may be used to avoid large scale oscillation problems.
N_HOLD	5 measurement periods	
T_RESPONSE	500 ms	Estimated response time for a request to be fulfilled.
Packet loss burst	2 or more packet losses in the last 20 packets.	

C.1.3.2 Adaptation state machine with four states

The first example utilizes all adaptation possibilities, both in terms of possible states and transitions between the states. Figure C.2 shows the layout of the adaptation state machine and the signalling used in the transitions between the states.

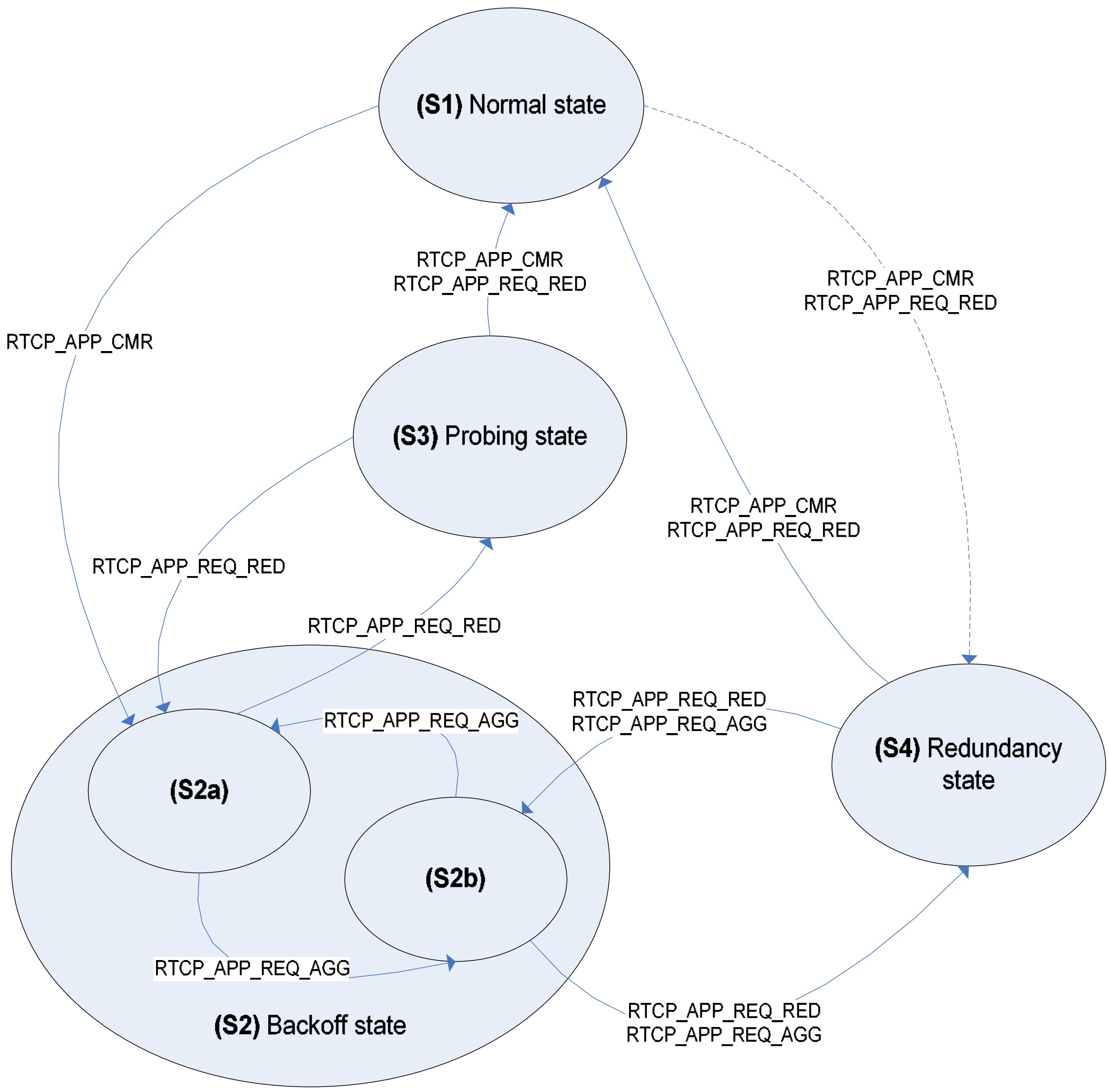


Figure C.2: State diagram for four-state adaptation state machine

State transitions:

Below are listed the possible state transitions and signalling that is involved. Note that the state can go from S1 to either S2 or state S4, this is explained below:

Table C.5: State transitions for four-state adaptation state machine

State transition	Conditions and actions
S1 → S2a	Condition: Packet loss ≥ PLR_1 or packet loss burst detected. A request to reduce the encoding bit-rate is sent using RTCP_APP_CMR, e.g. change mode from AMR 12.2 to AMR 5.9.
S2a → S2b	Condition: Packet loss ≥ PLR_1. This state transition occurs if the packet loss is still high despite the reduction in codec rate. A request is sent to reduce the packet rate is reduced by means of an RTCP_APP_REQ_AGG message.
S2b → S2a	Condition: Packet loss ≤ PLR_ 2 for N_HOLD consecutive measurement periods. This state transition involves an increase of the packet rate restoring it to the same value as in S1. The request transmitted is RTCP_APP_REQ_AGG. If the state transition S2b→S2a→S2b occurs in sequence, the state will be locked to S2b for N_INHIBIT frames to avoid state oscillation.
S2a → S3	Condition: Packet loss ≤ PLR_2 for N_HOLD consecutive measurement periods. A request to turn on 100% redundancy is transmitted by means of request RTCP_APP_REQ_RED.
S3 → S2a	Condition: Packet loss ≥ PLR_3. Same actions as in transition from, S1→S2a. If the transition S2a→S3→S2a→S3→S2a occurs, the S3 is disabled for N_INHIBIT frames.
S3 → S1	Condition: Packet loss ≤ PLR_2 for N_HOLD consecutive measurement periods. A request to turn off redundancy is transmitted as RTCP_APP _REQ_RED. Encoding bit-rate is increased by means of RTCP_APP_CMR.
S2b → S4	Condition: Packet loss ≥ PLR_3. A request to turn on 100% redundancy is transmitted by means of request RTCP_APP_REQ_RED. The packet rate is restored to same value as in S1 using RTCP_APP _REQ_AGG.
S4 → S2b	Condition: 1. If the previous transition was S2b→S4 and packet loss ≥ to 4*PLR@ S2b→S4 (packet loss considerably increased since transition to state S4). This is indicative of that the total bit-rate is too high and that it is probably better to transmit with a lower packet rate/bit-rate instead. This case might occur if the packet loss is high in S2a due to a congested link, a switch to redundant mode S4 will then increase the packet loss even more 2. If previous transition was S1→S4 and packet loss >= PLR_4. This transition is made to test if a bitrate/packet rate reduction is better.
S4 → S1	Condition: Packet loss < PLR_3 for N_HOLD consecutive measurement periods. A request to turn off redundancy is transmitted using RTCP_APP _REQ_RED. Encoding bit-rate is requested to increase using RTCP_APP_CMR.
S1 → S4	Condition: Packet loss ≥ PLR_1 or packet loss burst detected AND the previous transition was S4→S1, otherwise the transition S1→S2a will occur. A request to turn on 100% redundancy is transmitted using RTCP_APP_REQ_RED. The encoding bit-rate is requested to be reduced (in the example from AMR 12.2 to AMR 5.9) using RTCP_APP_CMR.

C.1.3.3 Adaptation state machine with four states (simplified version without frame aggregation)

This example is a simpler implementation with the frame aggregation removed.

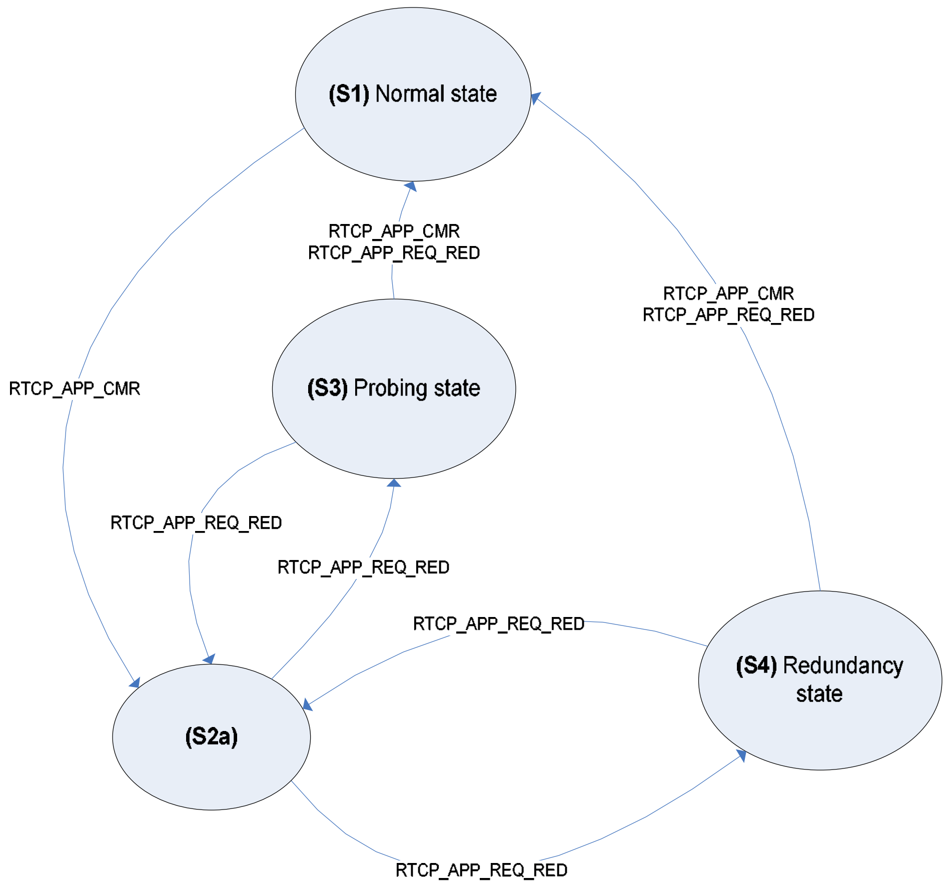


Figure C.3: State diagram for simplified four-state adaptation state machine

State transitions:

Below are listed the possible state transitions and signalling that is involved.

Table C.6: State transitions for simplified four-state adaptation state machine

State transition	Conditions and actions
S1 → S2a	Condition: Packet loss ≥ PLR_1 or packet loss burst detected. A request to reduce the encoding bit-rate is sent using RTCP_APP_CMR, e.g. change mode from AMR 12.2 to AMR 5.9.
S2a → S3	Condition: Packet loss ≤ PLR_2 for N_HOLD consecutive measurement periods. A request to turn on 100% redundancy is transmitted by means of request RTCP_APP_REQ_RED.
S3 → S2a	Condition: Packet loss ≥ PLR_3. Same actions as in transition from, S1→S2a. If the transition S2a→S3→S2a→S3→S2a happens in sequence state S3 is disabled for N_INHIBIT frames.
S3 → S1	Condition: Packet loss ≤ PLR_2 for N_HOLD consecutive measurement periods. A request to turn off redundancy is transmitted as RTCP_APP _REQ_RED. Encoding bit-rate is increased by means of RTCP_APP_CMR.
S2a → S4	Condition: Packet loss ≥ PLR_3. A request to turn on 100% redundancy is transmitted by means of request RTCP_APP_REQ_RED.
S4 → S2a	Condition: Packet loss ≥ to 4*PLR@ S2b→S4 (packet loss considerably increased since transition to state S4). This is indicative of that the total bit-rate is too high and that it is probably better to transmit with a lower packet rate/bit-rate instead. This case might occur if the packet loss is high in S2a due to a congested link, a switch to redundant mode S4 will then increase the packet loss even more.
S4 → S1	Condition: Packet loss < PLR_3 for N_HOLD consecutive measurement periods. A request to turn off redundancy is transmitted using RTCP_APP _REQ_RED. Encoding bit-rate is requested to increase using RTCP_APP_CMR.

C.1.3.4 Adaptation state machine with two states

This example is an implementation with the redundant states removed.

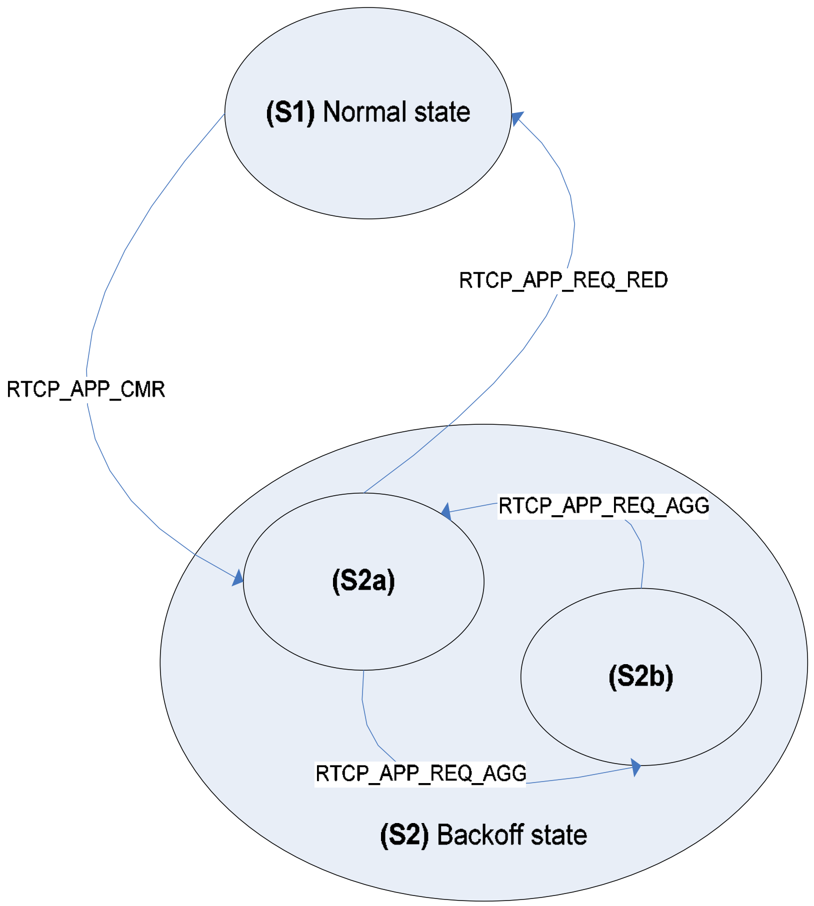


Figure C.4: State diagram for two-state adaptation state machine

State transitions:

Below are listed the possible state transitions and signalling that is involved.

Table C.7: State transitions for two-state adaptation state machine

State transition	Conditions and actions
S1 → S2a	Condition: Packet loss ≥ PLR_1 or packet loss burst detected. A request to reduce the encoding bit-rate is sent using RTCP_APP_CMR, e.g. change mode from AMR 12.2 to AMR 5.9. A failed transition counter counts the number of consecutive switching attempts S2a→S1 that fails. In the number of failed attempts is two or more state S1 is inhibited for N_INHIBIT frames. A failed transition attempt occurs if the previous transition was S2a→S1 and the state transition immediately occurs back to S2a.
S2a → S2b	Condition: Packet loss ≥ PLR_1. This state transition occurs if the packet loss is still high despite the reduction in codec rate. A request is sent to reduce the packet rate is reduced by means of an RTCP_APP_REQ_AGG message.
S2b → S2a	Condition: Packet loss ≤ PLR_2 for N_HOLD consecutive measurement periods. This state transition involves an increase of the packet rate. Also packet rate is restored to same value as in State (1) RTCP_APP_REQ_AGG. If the state transition S2b→S2a→S2b occurs in sequence, the state will be locked to S2b for N_INHIBIT frames.
S2a → S1	Condition: Packet loss ≤ PLR_2 for N_HOLD consecutive measurement periods. Redundancy is turned on (100 %) by means of request RTCP_APP_REQ_RED.

Annex D (informative): Reference delay computation algorithm

In this annex, the reference jitter management algorithm is described. It is written in pseudo code and is non-causal; hence non-implementable. The purpose of this algorithm is to define an "ideal" behaviour which all jitter buffers used in MTSI should strive to mimic. This buffer operates based on three input parameters:

- lookback factor to set the current target buffering depth;
- target late loss rate;
- maximum allowed time scaling percentage.

```
function ref_jb(channel,jb_adaptation_lookback,delay_delta_max,target_loss)
% channel         = file name of the channel
% lookback        = look back factor when estimating the max jitter
%                   buffer level [number of frames]
% delay_delta_max = max timescaling related modification (%) of the
%                   delay
% target_loss     = target late loss (%)
% example syntax:
% ref_jb('channel_1.dat',200,15,0.5);


framelength = 20;
% this value sets the speech data in each RTP packet to 20 ms. For 2 speech
% frames/RTP packet the value would be 40 ms.
jitter_est_window=50;
% Sets the jitter estimation window in number of frames
delay_delta_max_ms = framelength*delay_delta_max*0.01;
% Sets the maximum allowed time scaling
tscale = 1;
% Scale factor of delay data
% In this case the files are assumend to be ascii files with one delay
% entry per line, the entries are in ms, a negative value denotes
% a packet loss.
x = load(channel);
x =x';
% remove packet losses
% remove inital startup empty frames
ix = find(x > 0);
x(1:ix(1)-1) = x(ix(1));
% remove packet losses (replace with nearby delay values)
ix = find(x < 0);
packet_loss = length(ix)/length(x)*100;
for n=1:length(ix)
    if (ix(n) > 1)
        x(ix(n)) = x(ix(n)-1);
    end;
end;
% convert timescale to ms
x = x*tscale;
L = length(x);
T = 1:L;
% estimate min and max TX delay, estimate a delta_delay
for n=1:L
    ix = [max(1,n-jitter_est_window):n];
    max_delay(n) = max(x(ix));
    min_delay(n) = min(x(ix));
    delta_delay(n) = max_delay(n)-min_delay(n);
end
% compute the target max jitter buffer level with some slow adaptation
% downwards, just to mimick how a jitter buffer might behave
for n=1:L
    ix = [max(1,n-jb_adaptation_lookback):n];
    jb(n) = max(delta_delay(ix));
    % The timescaling is not allowed to adjust the jitterbuffer target max level
    % too fast.
    if n == 1
        jb_ = jb(n);
    end
    delta = abs(jb_-jb(n));
    if delta < delay_delta_max_ms;
```

```
        jb_ = jb(n);
    else
        if (jb(n) < jb_)
            jb_ = jb_-delay_delta_max_ms;
        else
            jb_ = jb_+delay_delta_max_ms;
        end
        jb(n) = jb_;
    end
    % jitter buffer target max level can only assume an integer number of frames
    jbq(n) = ceil(jb(n)/framelength)*framelength;
    % compute estimated delay
    del(n) = jbq(n)+min_delay(n);
end

if target_loss > 0
    % decrease the max jitter buffer leve until a target late loss has been
    % reached.
    late_loss = length(find(del < x))/L*100.0;
    jbq_save = jbq; % as the max level is increased until the late loss > target one
    % must be able to revert back to the previous data
    while late_loss < target_loss
        jbq_save = jbq;
        jbq = min(max(jbq)-framelength,jbq);
        del = jbq+min_delay;
        late_loss = length(find(del < x))/L*100.0;
    end
    jbq = jbq_save;
    del = jbq+min_delay;
end

jdel = max(0,del-x);
%Calculate and plot the CDF of the reference buffer.
figure(1);plot(T,jbq,T,del,T,x);
[n,x] = hist(jdel,140); y = cumsum(n);y = y/max(y)*100;
figure(2);plot(x,y);axis([0 200 0 100]);ylabel('%');xlabel('ms');title('CDF of packet delay in JB');
```

Annex E (informative): QoS profiles

E.1 General

This annex contains examples with mappings of SDP parameters to UMTS QoS parameters [64] for MTSI.

E.2 Bi-directional voice (AMR12.2 over IPv4, RTCP)

The bitrate for AMR 12.2 including IP overhead (one AMR frame per RTP packet, using bandwidth efficient mode) is 28.8 kbps which is rounded up to 29 kbps.

Table E.1: QoS mapping for bi-directional voice (AMR 12.2 over IPv4, RTCP)

Traffic class	Conversational class	Notes
Delivery order	No	The application should handle packet reordering.
Maximum SDU size (octets)	1 400	Maximum size of IP packets
Delivery of erroneous SDUs	No	
Residual BER	10^{-5}	Reflects the desire to have a medium level of protection to achieve an acceptable compromise between packet loss rate and voice transport delay and delay variation.
SDU error ratio	$7*10^{-3}$	A packet loss rate of 0.7 % per wireless link is in general sufficient for voice services
Transfer delay (ms)	130 ms	Indicates maximum delay for 95th percentile of the distribution of delay for all delivered SDUs between the UE and the GGSN during the lifetime of a bearer service. Permits the derivation of the RAN part of the total transfer delay for the UMTS bearer. This attribute allows RAN to set transport formats and H-ARQ/ARQ parameters such as the discard timer.
Guaranteed bit rate for uplink (kbps)	31	The bit-rate of AMR12.2 including IP/UDP/RTP overhead + 5 % for RTCP. This value applies for IPv4.
Maximum bitrate for uplink (kbps)	31	The same as the guaranteed bitrate.
Guaranteed bit rate for downlink (kbps)	31	The bit-rate of AMR12.2 including IP/UDP/RTP overhead + 5 % for RTCP. This value applies for IPv4.
Maximum bitrate for downlink (kbps)	31	The same as the guaranteed bitrate
Allocation/Retention priority	subscribed value	Indicates the relative importance to other UMTS bearers. It should be the next lower value to the priority of the signalling bearer.
Source statistics descriptor	‘speech'	

E.3 Bi-directional video (128 kbps, IPv4 and RTCP)

The video bandwidth is assumed to be 120 kbps and the IP overhead 8 kbps, resulting in 128 kbps. The transfer delay for video is different from other media.

Table E.2: QoS mapping for bi-directional video (128 kbps, IPv4, RTCP)

Traffic class	Conversational class	Notes
Delivery order	No	The application should handle packet reordering.
Maximum SDU size (octets)	1 400	Maximum size of IP packets
Delivery of erroneous SDUs	No	
Residual BER	10^{-5}	Reflects the desire to have a medium level of protection to achieve an acceptable compromise between packet loss rate and voice transport delay and delay variation.
SDU error ratio	$7*10^{-3}$	A packet loss rate of 0.7 % per wireless link is in general sufficient for video services
Transfer delay (ms)	170 ms	Indicates maximum delay for 95th percentile of the distribution of delay for all delivered SDUs between the UE and the GGSN during the lifetime of a bearer service. Permits the derivation of the RAN part of the total transfer delay for the UMTS bearer. This attribute allows RAN to set transport formats and H-ARQ/ARQ parameters such as the discard timer.
Guaranteed bit rate for downlink (kbps)	144	The bit-rate of a video codec running at 128 kbps including IP/UDP/RTP overhead (assumed to be 8 kbps) and RTCP (adds 5 %) rounded up to nearest 8kbps value. This value applies for IPv4.
Maximum bit rate for downlink (kbps)	144	The same as the guaranteed bitrate.
Guaranteed bit rate for uplink (kbps)	144	The bit-rate of a video codec running at 128 kbps including IP/UDP/RTP overhead (assumed to be 8 kbps) and RTCP (adds 5%) rounded up to nearest 8 kbps value. This value applies for IPv4.
Maximum bitrate for uplink (kbps)	144	The same as the guaranteed bitrate.
Allocation/Retention priority	subscribed value	Indicates the relative importance to other UMTS bearers. It should be the same or next lower value to the priority of a Conversational bearer with source statistics descriptor 'speech'.
Source statistics descriptor	'unknown'	

E.4 Bi-directional real-time text (3 kbps, IPv4, RTCP)

Bi-directional text at 3 kbps all inclusive (text, IP overhead, RTCP).

Table E.3: QoS mapping for bi-directional real-time text (3 kbps, IPv4, RTCP)

Traffic class	Conversational class	Notes
Delivery order	No	The application should handle packet reordering.
Maximum SDU size (octets)	1 400	Maximum size of IP packets
Delivery of erroneous SDUs	No	
Residual BER	10^{-5}	Reflects the desire to have a medium level of protection to achieve an acceptable compromise between packet loss rate and voice transport delay and delay variation.
SDU error ratio	$1*10^{-3}$	Text should have a higher level of protection than voice and video.
Transfer delay (ms)	130 ms	Indicates maximum delay for 95th percentile of the distribution of delay for all delivered SDUs between the UE and the GGSN during the lifetime of a bearer service. Permits the derivation of the RAN part of the total transfer delay for the UMTS bearer. This attribute allows RAN to set transport formats and H-ARQ/ARQ parameters such as the discard timer.
Guaranteed bit rate (kbps)	3.0	An assumed bit-rate of a real-time text service including headers and RTCP.
Maximum bitrate (kbps)	3.0	The same as the guaranteed bitrate.
Guaranteed bit rate (kbps)	3.0	An assumed bit-rate of a real-time text service including headers and RTCP.
Maximum bitrate (kbps)	3.0	The same as the guaranteed bitrate.
Allocation/Retention priority	Subscribed value	Indicates the relative importance to other UMTS bearers. It should be a lower value to the priority of a Conversational bearer with source statistics descriptor ‘speech'.
Source statistics descriptor	‘unknown'	

Annex F (normative): Media types, codecs and formats used for MSRP transport in MTSI

F.1 General

The IMS messaging service is described in TS 26.141 [59]. The description of IMS messaging in clauses 1-6 of 3GPP TS 26.141 [59] is applicable for MSRP-transported media in MTSI. The MSRP transport itself is described in 3GPP TS 24.173 [57].

All statements in TS 26.141 regarding IMS messaging are valid for MSRP transported media in MTSI including the status of the statement (shall, should, may).

Any differences between IMS messaging in 3GPP TS 26.141 [59] and MSRP transported media in MTSI are described in clause F.2.

F.2 Difference relative to 3GPP TS 26.141

F.2.1 Video

For MSRP transported Media in MTSI, clause 5.3 in 3GPP TS 26.141 [59] is void and instead the following shall be used.

If an MSRP client supports video, ITU-T Recommendation H.263 profile 0 Level 45 decoder [22] shall be supported. In addition, an MSRP client should support:

- H.263 Profile 3 Level 45 decoder [22];
- MPEG-4 Visual Simple Profile Level 3 decoder [23] with the following constraints:
  - Number of Visual Objects supported shall be limited to 1.
  - The maximum frame rate shall be 30 frames per second.
  - The maximum f_code shall be 2.
  - The intra_dc_vlc_threshold shall be 0.
  - The maximum horizontal luminance pixel resolution shall be 352 pels/line.
  - The maximum vertical luminance pixel resolution shall be 288 pels/VOP.
  - If AC prediction is used, the following restriction applies: QP value shall not be changed within a VOP (or within a video packet if video packets are used in a VOP). If AC prediction is not used, there are no restrictions to changing QP value.
- H.264 (AVC) Baseline Profile Level 1b decoder [24] with constraint_set1_flag=1.1 and without requirements on output timing conformance (Annex C of [24]).

The video buffer model given in Annex G of document [60] should be supported if H.263 or MPEG-4 Visual is supported. It shall not be used with H.264 (AVC).

NOTE: ITU-T Recommendation H.263 profile 0 has been mandated to ensure that video-enabled MSRP clients support a minimum baseline video capability. Both H.263 and MPEG-4 Visual decoders can decode an H.263 profile 0 bitstream. It is strongly recommended, though, that an H.263 profile 0 bitstream is transported and stored as H.263 and not as MPEG-4 Visual (short header), as MPEG-4 Visual is not mandated by MTSI.

Annex G (Normative): DTMF events

G.1 General

This annex describes a method for sending DTMF events in the same RTP media stream as the speech.

- MTSI terminals offering speech communication shall support the below described method in the transmitting direction and should support it in the receiving direction.
- MTSI media gateways offering speech communication shall support the below described method in both the transmitting and the receiving direction. For MTSI media gateways, the described method applies only to the PS session between the gateway and an MTSI terminal.

This method was designed to send DTMF events in the same RTP streams as the speech.

G.2 Encoding of DTMF signals

DTMF should be encoded and transmitted as DTMF events. DTMF events in this Annex refers to the DTMF named events described in Section 3.2, Table 3 in [61], i.e. events (0-9,A-D, *, #) which are encoded with event codes 0—9, 10, 11 and 12 – 15 respectively. DTMF events can either be narrowband or wideband, i.e. use 8 kHz or 16 kHz sampling frequency respectively. MTSI terminals and media gateways that support both narrowband and wideband speech shall support both narrowband and wideband DTMF events. When switching between speech and DTMF, the DTMF events should use the same sampling frequency as for the speech that is currently being transmitted.

The encoding of DTMF events includes specifying the duration time for the events, [61]. To harmonize with legacy DTMF signalling, [62], [63], the tone duration of a DTMF event shall be at least 65 ms and the pause duration in-between two DTMF events shall be at least 65 ms. The duration of the DTMF event and the pause time to the next DTMF event, where applicable, should be selected such that it enables incrementing RTP Time Stamp with a multiple of the number of timestamp units corresponding to the frame length of the speech codec used for the speech media.

G.3 Session setup

An MTSI terminal or media gateways offering a speech media session for speech and DTMF events should include an offer for DTMF events according to the example in Table G.3.1 when narrowband speech is offered and according to the example in Table G.3.2 when both narrowband and wideband speech is offered. The answerer shall select DTMF payload format(s) that match the selected speech codec(s).

Table G.3.1: SDP example for narrowband speech and DTMF

SDP offer
m=audio 49152 RTP/AVP 97 98 99
a=rtpmap:97 AMR/8000/1
a=fmtp:97 mode-change-capability=2; max-red=220
a=rtpmap:98 AMR/8000/1
a=fmtp:98 mode-change-capability=2; max-red=220; octet-align=1
a=rtpmap:99 telephone-event/8000/1
a=fmtp:99 0-15
a=ptime:20
a=maxptime:240

Table G.3.2: SDP example narrowband and wideband for both speech and DTMF

SDP offer
m=audio 49152 RTP/AVP 97 98 99 100 101 102
a=rtpmap:97 AMR-WB/16000/1
a=fmtp:97 mode-change-capability=2; max-red=220
a=rtpmap:98 AMR-WB/16000/1
a=fmtp:98 mode-change-capability=2; max-red=220; octet-align=1
a=rtpmap:99 telephone-event/16000/1
a=fmtp:99 0-15
a=rtpmap:100 AMR/8000/1
a=fmtp:100 mode-change-capability=2; max-red=220
a=rtpmap:101 AMR/8000/1
a=fmtp:101 mode-change-capability=2; max-red=220; octet-align=1
a=rtpmap:102 telephone-event/8000/1
a=fmtp:102 0-15
a=ptime:20
a=maxptime:240

G.4 Data transport

When sending and receiving DTMF events with RTP the RTP payload format for DTMF digits, telephony tones, and telephony signals, RFC 4733 [61], shall be supported.

DTMF events shall use the same media stream as for speech, i.e. the same IP number, UDP port and RTP SSRC. Thereby, RTP Sequence Number and RTP Time Stamp shall be synchronized between speech and DTMF. For example, by setting the initial random values the same and when switching from speech to DTMF, or vice versa, the RTP Sequence Number and RTP Time Stamp shall continue from the value that was used for the other audio media (speech or media).

The RTP Sequence Number shall increment in the same way as for speech, i.e. by 1 for each transmitted packet.

The RTP Time Stamp should increment in the same way as for speech packets or with a multiple, i.e. if the RTP Time Stamp increments with 160 between speech packets then the increment for DTMF should be 160 or a multiple of 160. The RTP Time Stamp should not increment with a smaller interval for DTMF than for speech. The RTP Time Stamp should use the same sampling frequency as for the speech that is transmitted immediately before the start of the DTMF event(s).

NOTE: One DTMF event may be transmitted in several RTP packets, for example if the event is a long-lasting event. In this case all RTP packets containing the same DTMF event shall have the same RTP Time Stamp value according to RFC 4733 [61].

Speech packets shall not be transmitted when DTMF events are transmitted in the same RTP media stream.

Annex H (informative): Change history

Change history							
Date	**TSG #**	**TSG Doc.**	**CR**	**Rev**	**Subject/Comment**	**Old**	**New**
2007-03	35	SP-070110			Approved at TSG SA #35	2.0.0	7.0.0
2007-06	36	SP-070318	0001	2	Addition of non-compound RTCP	7.0.0	7.1.0
2007-06	36	SP-070318	0002	1	Addition of DTMF	7.0.0	7.1.0
2007-06	36	SP-070318	0005	2	Video QoS Profile	7.0.0	7.1.0
2007-06	36	SP-070318	0006	1	Correction of the reference to the AMR/AMR-WB RTP payload format	7.0.0	7.1.0
2007-06	36	SP-070318	0007	1	Video rate adaptation in MTSI	7.0.0	7.1.0
2007-06	36	SP-070318	0008	1	Improved Video support for MTSI	7.0.0	7.1.0

Rapporteur:

Per Fröjdh, Ericsson <per.frojdh@ericsson.com>

History

Document history		
V7.1.0	June 2007	Publication