ETSI TR 126 914 V7.0.0 (2007-06)

Technical Report

Digital cellular telecommunications system (Phase 2+);
Universal Mobile Telecommunications System (UMTS);
Multimedia telephony over IP Multimedia Subsystem (IMS);
Optimization opportunities
(3GPP TR 26.914 version 7.0.0 Release 7)

GLOBAL SYSTEM FOR
MOBILE COMMUNICATIONS

Reference
DTR/TSGS-0426914v700

Keywords
GSM, 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

Important notice

Individual copies of the present document can be downloaded from:
http://www.etsi.org

The present document may be made available in more than one electronic version or in print. In any case of existing or perceived difference in contents between such versions, the reference version is the Portable Document Format (PDF). In case of dispute, the reference shall be the printing on ETSI printers of the PDF version kept on a specific network drive within ETSI Secretariat.

Users of the present document should be aware that the document may be subject to revision or change of status. Information on the current status of this and other ETSI documents is available at
http://portal.etsi.org/tb/status/status.asp

If you find errors in the present document, please send your comment to one of the following services:
http://portal.etsi.org/chaircor/ETSI_support.asp

Copyright Notification

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 Report (TR) 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 Report 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 over IP (3GPP TS 22.973 [2]) 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.

While the user experience of Multimedia telephony is expected to have some similarity to existing telephony services, the richer capabilities of IMS is exploited. In particular, multiple media components, such as voice and video, can be used and dynamically added or dropped during a call.

1   Scope

The present document:

- identifies opportunities for optimization of service quality and efficiency of Multimedia telephony over IP in a qualitative sense;

- provides the basis for developing a set of optional backward-compatible tools implementing such optimizations.

The optimized multimedia telephony targets many different system configurations and operating conditions, e.g. GERAN, UTRAN, inter-working between GERAN, UTRAN, GAN, and different PLMNs. Of the various use cases for multimedia telephony, the main focus of the present document is on voice calls, but the focus also includes other core media components, such as video. In particular, areas with optimization opportunities include handling of degraded channels, delay jitter, packet losses, efficiency, inter-working with other voice systems, etc.

The scope includes proposing solutions that maintain backward compatibility in order to ensure seamless inter-working with existing services available in the CS domain, such as CS voice telephony and 3G-324M, as well as with terminals of earlier 3GPP releases. Alignment with legacy media formats avoids transcoding and even allows realizing at least parts of the optimization gains. All optimizations are hence based on the default codecs specified in 3GPP TS 26.235 [3] and 3GPP TS 26.141 [6].

The optimizations identified in the present document address mainly. media transport and signalling. Most of the SIP signalling is out of scope and is handled by other 3GPP groups. Issues regarding registration to the network and/or to IMS at power-on or at other occurrences are not included either.

The optimization tools are not specified in the present document per se, but will be specified as amendments to existing (pre-Release 7) Technical Specifications (3GPP TS 26.235 [3] and 3GPP TS 26.236 [4]) and possibly new Technical Specifications. Furthermore, a characterization of the optimized multimedia telephony over IMS will be available in a separate Technical Report.

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 TR 22.973: "IMS Multimedia Telephony Communication Enabler and supplementary services".

[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 26.935: "Packet Switched (PS) conversational multimedia applications; Performance characterization of default codecs".

[6]   3GPP TS 26.141: "IP Multimedia System (IMS) Messaging and Presence; Media formats and codecs".

[7]   3GPP TS 43.318: "Generic access to the A/Gb interface; Stage 2".

[8]   3GPP TR 45.912: "Feasibility study for evolved GSM/EDGE Radio Access Network (GERAN)".

[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]   IETF RFC 2327 (1998): "SDP: Session Description Protocol", M. Handley M. and V. Jacobson.

[12]   3GPP TS 26.071: "AMR Speech Codec; General description".

[13]   3GPP TS 26.090: "AMR Speech Codec; Transcoding functions".

[14]   3GPP TS 26.073: "AMR Speech Codec; C-source code".

[15]   3GPP TS 26.104: "ANSI-C code for the floating-point Adaptive Multi-Rate (AMR) speech codec".

[16]   3GPP TS 26.171: "AMR Speech Codec, wideband; General description".

[17]   3GPP TS 26.190: "Speech codec speech processing functions; Adaptive Multi-Rate - Wideband (AMR-WB) speech codec; Transcoding functions".

[18]   3GPP TS 26.173: "ANCI-C code for the Adaptive Multi Rate - Wideband (AMR-WB) speech codec".

[19]   3GPP TS 26.204: "ANSI-C code for the floating-point Adaptive Multi-Rate Wideband (AMR-WB) speech codec".

[20]   IETF RFC 3267 (2002): "Real-Time Transport Protocol (RTP) Payload Format and File Storage Format for the Adaptive Multi-Rate (AMR) Adaptive Multi-Rate Wideband (AMR-WB) Audio Codecs", J. Sjoberg, M. Westerlund, A. Lakaniemi and Q. Xie.

[21]   ITU-T Recommendation H.263 (1998): "Video coding for low bit rate communication".

[22]   ITU-T Recommendation H.263 - Annex X (2004): "Profiles and levels definition".

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

[27]   IETF RFC 3095 (2001): "RObust Header Compression (ROHC): Framework and four profiles: RTP, UDP, ESP, and uncompressed", C. Bormann and al.

[28]   3GPP TS 26.234: "Transparent end-to-end Packet-switched Streaming Service (PSS); Protocols and codecs".

3   Definitions and abbreviations

3.1   Definitions

For the purposes of the present document, the following terms and definitions apply:

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:

ARQ	Automatic repeat ReQuest
AS	Application Server
AVC	Advanced Video Coding
CSCF	Call Session Control Function
GAN	Generic Access Network
H-ARQ	Hybrid - ARQ
HSDPA	High Speed Downlink Packet Access
IMS	IP Multimedia Subsystem
IP	Internet Protocol
IPv4	Internet Protocol version 4
IPv6	Internet Protocol version 6
ITU-T	International Telecommunications Union - Telecommunications
MM	MultiMedia
MMS	Multimedia Messaging Service
QoS	Quality of Service
RTP	Real-time Transport Protocol
SDP	Session Description Protocol
SIP	Session Initiation Protocol
TBF	Temporary Block Flow
ToIP	Telephony over IP
TTI	Transmission Time Interval
URL	Universal Resource Locator
VoIP	Voice over IP

4 Overview

4.1 General

The rest of the present document is divided into five main clauses.

Clause 4 contains an architectural overview of the IMS system indicating the domain relevant to the work reported in the present document.

Clause 5 provides the assumptions underlying the work of this report. General assumptions refer to networks, codecs and backward compatibility, etc. The following subclauses give information relevant to the channels, media components as well as use scenarios for multimedia telephony that are relevant to the optimizations targeted by the present document.

Clause 6 shows the current status in Release 6 for non-optimized multimedia telephony. The point-to-point voice call has been used as the basic "use case". The goal for this use case is to show what is needed to create a VoIP telephony service with the same quality as circuit-switched telephony when the channel is virtually error-free. The following subclause also includes the current specifications for using video in multimedia telephony.

Clause 7 is a list of areas with optimization opportunities in Release 7. Each area, such as handling of delay jitter, packet losses, inter-working with CS, inter-working between GERAN and UTRAN, etc., is a candidate for optimization where suggestions for standardization will be considered.

Clause 8 finally gives a summary of all findings of clause 7 as well as recommendations for further standardization efforts.

4.2 System description
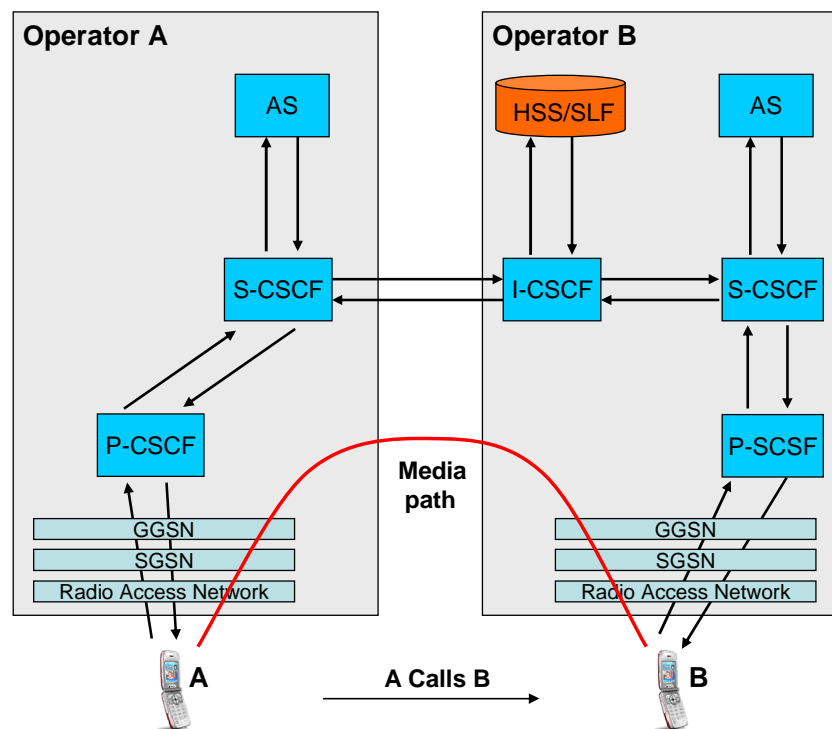
**Figure 1: High-level architecture figure showing the nodes involved
in a Multimedia Telephony call set-up over IMS**

A Multimedia Telephony call over IMS uses the Call Session Control Function (CSCF) mechanisms to route control-plane signalling between the UEs involved in the call (see figure 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 optimize the media path. In the example in figure 1, it is routed directly between the GGSNs outside the IMS. In case of trans-coding, multi-party calls or teleconferencing, the media path may pass through a Media Resource Function (MRF) that could handle mixing and transcoding of media streams.

5 Assumptions

5.1 General assumptions

The following general assumptions are used in the present document:

- The multimedia telephony service must be capable of operating on very different system configurations and operating conditions, for example GERAN, UTRAN, in combination with GAN (3GPP TS 43.318 [7]) and low/high system load.

- Backward compatibility (TFO and/or TrFO) with legacy systems (circuit switched GERAN and UTRAN) is required, for example by avoiding transcoding.

- Current codecs, payload formats, and session-signalling protocols standardized in 3GPP specifications will be used when applicable.

- If adaptation is used to handle congestion, then the UE shall not adapt so that the bit rate is significantly increased. The reason is that it is not a good idea to send more bits or packets into an already congested network.

- Backward compatibility is required to ensure seamless inter-working with existing services available in the CS domain, such as CS voice telephony and 3G-324M, as well as with terminals of earlier 3GPP releases.

- Terminal performance shall be the same as for corresponding CS voice services and follow the same guidelines. To this end, echo cancellation and noise suppression make significant contributions to ensure end-user quality of voice calls.

5.2   Channels

Multimedia telephony over IP is an access-independent service. Based on IP transport, it is possible to use on any IP-based network, although certain access types will put limitations on, e.g. available bit-rate and latency. However, there is a large span in characteristics among the possible access types and since the present document aims to show possible optimizations, there is a need to clarify under what conditions these proposed optimizations will benefit the service the most. Hence, the proposed optimizations are derived assuming 3GPP standardized access types, both in GERAN and UTRAN networks.

There are many uncertainties about the system configuration and operating conditions under which the multimedia telephony services will operate, especially when it comes to channel conditions. In 3GPP networks there are basically two different categories of radio channels; dedicated channels and shared channels. Dedicated channels allocate resources for a specific service for a specific user at the session set-up and are therefore more predictable when it comes to channel characteristics. Shared channels on the other hand will experience the same radio conditions as the dedicated channels, but the very nature of shared channels also introduces traffic load dependencies already in the scheduler. Hence, the traffic in channels shared between different users experiences not only varying radio conditions due to the current physical environment, but also a competition for transmission capacity between different users. In a shared channel, the channel scheduler has to make prioritizations between how to schedule the incoming packets to the shared channel in the cell for transmission. The scheduler can be optimized in different ways, but since the actual scheduling algorithms are proprietary, it is impossible on a scheduler-agnostic level to know exactly how packets under a certain network load for a certain service will be scheduled for transmission. Hence, although it may not possible to say what 'typical' shared channel characteristics look like, it is clear that from an application point-of-view they will be experienced as variable delay and packet losses.

HSDPA is an access type standardized in 3GPP Release 5. It is by its very nature one of the possible 3GPP access types that multimedia telephony will use. Although no general channel characteristics are applicable for an HSDPA radio link, two features are well known. First, HSDPA has fast H-ARQ re-transmissions on the radio layer, which makes packet losses due to bad radio conditions rather infrequent. Second, the scheduler needs to prioritize between different users and different services within one cell, which as a consequence will induce varying transport delay to the UE. Since the radio access network will induce network load dependencies in the transport characteristics, in combination with the proprietary scheduler and the lower layer re-transmissions, there is a clear need for application-layer handling of these varying characteristics, i.e. delay jitter and packet losses induced by late delivery. Packets can be lost not only due to late delivery in the UE but also due to other circumstances such as too many H-ARQ re-transmissions, flushed NodeB buffers at handover, etc.

In commercial 3GPP networks deployed today, dedicated channels are used for all media transport. Their characteristics are well known and they do not show the same behaviour as shared channels when it comes to delay jitter and packet loss. Dedicated channels may also be used for multimedia telephony so any optimizations suggested in this report must be compatible with dedicated channel characteristics, even if they in some cases are designed to handle induced jitter and packet loss most prominent in HSDPA.

There is currently an ongoing feasibility study (3GPP TR 45.912 [8]) in GERAN groups to analyze how GERAN may be evolved. This work includes, for example, shorter TTI, fast retransmissions and pre-allocated TBF, in order to, for example, increase peak bit rate, reduce latency and reduced end-to-end round-trip-time.

5.3   Media components

Multimedia telephony supports simultaneous transfer of multiple media components with real-time characteristics. Media components denote the actual components that the end-user experiences. Any two such components may or may not use the same underlying protocols.

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

- **Voice:** The sound that is picked up by a microphone and transferred from terminal A to terminal B and played out in a loudspeaker. In this report it is assumed that the call is full-duplex voice, although in principle it could be a one-way flow.

- **Video:** The moving image that is captured by a camera of terminal A and rendered on the display of terminal B. Video can be full or half duplex.

The present document focuses on core media components that 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.

5.4 Use scenarios

The following example use scenarios may benefit from the optimizations identified by the present document.

- Basic point-to-point voice telephony equivalent to today's CS voice service in UTRAN and GERAN. The main difference from CS is that the transport layers are replaced with IP.

- Basic point-to-point video telephony equivalent to today's CS video telephony in UTRAN. Both the voice and the (full duplex) video are enabled from call setup to call termination.

- Enriched point-to-point voice telephony where the video component is added and removed during the session. Voice is enabled the whole time, from call setup to call termination. Video (one or two way) may be enabled and disabled multiple times during the session.

Multimedia telephony can include several other types of use cases, but the above use cases involve the core components voice and video, which benefit most from the proposed optimizations in the present document.

6 Basic point-to-point call in Release 6

NOTE: This clause (6) only reviews technology already present in the Release-6 versions of 3GPP TS 26.235 [3] and 3GPP TS 26.236 [4].

6.1 General

This clause describes basic (Release 6) packet-switched conversational multimedia calls in terms of a point-to-point voice call and a point-to-point video call. The main focus is to describe how a session is set up in order to get a quality that is equivalent to CS point-to-point calls using TFO/TrFO and to show the media flow. It is assumed that the call is between two subscribers within the same PLMN. A characterization of the current voice codecs can be found in 3GPP TR 26.935 [5].

6.2   Session setup

The following example shows how a subscriber invites another subscriber to a VoIP session by issuing a SIP INVITE command with an SDP (RFC 2327 [11]) description of the session. In order to highlight the SDP part, it is shown in **bold**.

EXAMPLE 1:   INVITE tel:+1-212-555-2222 SIP/2.0
Via: SIP/2.0/UDP [5555::aaa:bbb:ccc:ddd]:1357;comp=sigcomp;branch=z9hG4bKnashds7
Max-Forwards: 70
Route: <sip:pcscf1.visited1.net:7531;lr;comp=sigcomp>, <sip:scscf1.home1.net;lr>
P-Preferred-Identity: "John Doe" <sip:user1@home1.net>
Privacy: none
From: <sip:user1@home1.net>;tag=171828
To: <tel:+1-212-555-2222>
Call-ID: cb03a0s09a2sdfglkj490333
Cseq: 127 INVITE
Contact: <sip:[5555::aaa:bbb:ccc:ddd]:1357;comp=sigcomp>
Allow: INVITE, ACK, CANCEL, BYE, PRACK, UPDATE, REFER, MESSAGE
Content-Type: application/sdp
Content-Length: (222)

v=0
o=- 2987933615 2987933615 IN IP6 5555::aaa:bbb:ccc:ddd
s=-
c=IN IP6 5555::aaa:bbb:ccc:ddd
t=0 0
m=audio 3456 RTP/AVP 97
b=AS:16
a=rtpmap:97 AMR/8000
a=fmtp:97 mode-change-period=2
a=maxptime:20
a=ptime:20

An offer to use wide-band speech may be realized by replacing the media part of the above SDP with the following lines:

EXAMPLE 2:   **m=audio 49120 RTP/AVP 98**
a=rtpmap:98 AMR-WB/16000
a=fmtp:98 mode-change-period=2
a=maxptime:20
a=ptime:20

An offer to use H.263 video may be realized by adding the following media part to the SDP:

EXAMPLE 3:   **m=video 51372 RTP/AVP 99**
a=rtpmap:99 H263-2000/90000
a=fmtp:99 profile=0;level=45

6.3   Media flow

6.3.1   General

The user plane protocol stack for a multimedia telephony capable terminal is shown below. All media components are transported over RTP with each respective payload format mapped onto the RTP (RFC 3550 [9]) flow according to RFC 3551 [10].

Conversational Multimedia Application			
Audio	Video	Text	RTCP
Payload formats			RTCP
RTP			RTCP
UDP			
IP			
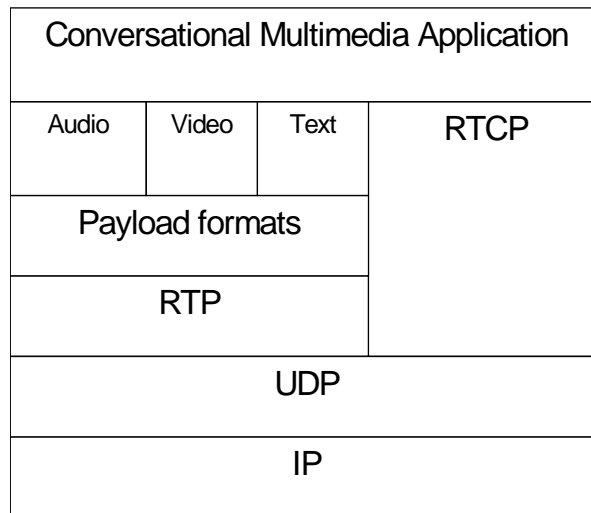
Figure 2: User plane protocol stack for a multimedia telephony capable terminal

The information in the RTP header, together with the payload header, is used for different purposes in the call.

- The sequence number is used for detection of lost packets.

- SSRC is used for source tracking. For voice calls a typical use would be in tele-conferencing.

6.3.2　Voice

For a point-to-point voice call, the default speech codec is AMR (3GPP TS 26.071 [12], 3GPP TS 26.090 [13], 3GPP TS 26.073 [14] and 3GPP TS 26.104 [15]), or in case of wideband speech, AMR-WB (3GPP TS 26.171 [16], 3GPP TS 26.190 [17], 3GPP TS 26.173 [18] and 3GPP TS 26.204 [19]). Both codecs use the payload formats specified in RFC 3267 [20]. There are a number of configurable details available in the payload format settings. The PSC specification for transport protocols (3GPP TS 26.236 [4]) states the following settings for AMR, AMR-WB encoded speech packetized according to RFC 3267 [20] for use in a Release 6 point-to-point packet switched multimedia voice call.

Sender behaviour:

- The bandwidth efficient operation shall be used. The alternative to bandwidth efficient operation is octet-aligned operation. The former reduces overhead while the latter enables the use of frame interleaving and other robustification procedures (see RFC 3267 [20] for details).

- Codec mode changes shall be performed in integer multiples of 40 ms. This is for backward compatibility with CS GSM where only codec mode indication *or* codec mode request could be sent in one frame.

- Codec mode changes should be performed to neighbouring modes of the selected combination of codec modes.

- DTX signalling may be used.

- Only one speech frame shall be encapsulated in each RTP packet. While this provides a big IP overhead for the media flow, it keeps the transport-induced latency low. The IP overhead can be significantly reduced by using header compression (RFC 3095 [27]).

- The multi-channel session shall not be used. Only mono streams are allowed even though the payload format allows for multi-channel sessions.

- Interleaving shall not be used. Although robustness is increased, the latency is also significantly increased which will substantially reduce conversational quality.

- Internal CRC shall not be used.

Receiver behaviour:

- Codec mode changes shall be accepted at any time.

- Codec mode changes shall be accepted to any supported mode of the selected combination of codec modes.

- DTX signalling shall always be accepted.

The information in the RTP header, together with the payload header, is used for different purposes in the call.

- The marker bit is used when the frame contains the first speech frame in a talk spurt (i.e. first frame after DTX has been used). The marker bit identifies suitable locations in the media stream where buffer adaptation is particularly useful.

- The timestamp field should be used for frame sorting in the receiver buffer and detection of lost frames.

- The timestamp field is incremented according to the sampling frequency of the codec specified in the SDP (8 kHz for AMR-NB and 16 kHz for AMR-WB). This field is increased also in DTX mode (eight times the length of the typical frame increment) and sent in the SID UPDATE packet.

6.3.3 Video

Packet-switched multimedia terminals offering video communication shall support ITU-T Recommendation H.263 [21] and [22] Profile 0 Level 45. In addition they should support ITU-T Recommendation H.263 [21] and [22] Profile 3 Level 45, MPEG-4 Visual Simple Profile Level 0b (ISO/IEC 14496-2 [23]), and H.264 (AVC) Baseline Profile Level 1b (ITU-T Recommendation H.264 [24]) without requirements on output timing conformance (Annex C of ITU-T Recommendation H.264 [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 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: Terminals may use full-frame freeze and full-frame freeze release SEI messages of H.264 (AVC) to control the display process.

NOTE 2: 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.

Video packets should not be large to allow better error resilience and to minimize the transmission delay in conversational service. The size of each packet shall be kept smaller than 512 bytes.

7 Areas of optimizations

7.1 General

This clause identifies areas of multimedia telephony over IMS where opportunities for optimization exist. A number of problems (potential optimization areas) are identified in the subclauses below.

7.2   Delay jitter handling

Shared channels in conjunction with packet-based transport are well known to introduce variations in end-user packet receiver timing. This is known as delay jitter. For real-time services, delay jitter poses a significant threat to the service quality and, in the conversational case, the conversational quality. The source of jitter can vary; for shared channels the jitter might occur due to high load in the cell where the scheduler has many different streams to schedule for transmission, due to the current service mix in the cell having different streams with different transmission requirements or due to other circumstances. Channels can have other jitter sources such as lower layer re-transmissions. The lower layer fast re-transmissions available in HSDPA will push the packet loss due to bad radio conditions to a minimum but it will do so at the expense of varying delay. Since HSDPA does not support soft handover, delay variations might also occur during handover. Although no shared channels for media transport is available in GERAN, features like acknowledged radio links can give similar consequences with varying packet delivery timing in the UE.

The presence of delay jitter is not by itself a threat to the real-time service performance; it is the actual magnitude of the jitter that will determine the service quality impact.

A common solution to handle delay jitter is by introducing a jitter buffer. This buffer is designed to handle the variation in packet receiver timing. However, the mere presence of the buffer implicates that the service latency will be increased. In order to minimize the additional latency introduced by the jitter buffer, various adaptive schemes have been proposed. The main goal of the schemes is to minimize the delay while at the same time prevent packet losses due to packet delivery timing variations.

For multimedia telephony, delay jitter is a problem that needs to be addressed. This new functionality, jitter handling, not previously mentioned in any official 3GPP specification, is a pre-requisite for the performance of the real-time critical components in multimedia telephony over IMS with the voice component as the most obvious example.

7.3   Packet-loss handling

Clause 5 of the present document mentions that the multimedia telephony service must be capable of operating on several different system configurations and also at many different load levels. Even though significant development is still in progress, one can still draw some conclusions for the operating conditions that can be expected:

- For UTRAN DCH channels, where the fast power control tries to keep the block error rate at a fairly constant level, one should expect that a few percent of packet loss may occur.

- For UTRAN HSDPA/HSUPA the MAC-layer fast re-transmission scheme will usually try to keep the packet loss rate very close to zero. However, due to the scheduling, which is vendor specific, one cannot guarantee that the packet loss rate will be this low for all operating conditions. Depending on the load level, one can expect temporary packet loss rates that are at least as high as for the DCH channels, or even higher.

  NOTE: The performance will depend on the final specification of HSDPA/HSUPA RABs, e.g. whether RLC Acked or Un-acked mode is used.

- For GERAN, where significant development is being made, it is very unsure what the performance one can expect. For current GERAN (Release 6) the RLC retransmission time will be far too slow to be useful for real-time services with delay constraints that these services require. Therefore, RLC Un-acked Mode will have to be used, which means that the service must be capable of handing packet losses, at least up to a few percent.

Some of the media codecs specified in Release 6 of 3GPP TS 26.235 [3] do not address decoder reaction to non-compliant bitstreams resulting from, e.g. packet loss. Investigate and provide appropriate solutions to maintain acceptable media quality in typical 3GPP environments. A feedback mechanism to report error-events may be useful in order to mitigate the effects of packet loss and maintain the quality in error-prone 3GPP environments.

7.4   Handling of inter-working with CS

There is a list of issues concerning inter-working with CS. For voice this comprises questions related to facilitating proper inter-working with the speech codecs defined in 3GPP TS 26.103 [26]. In particular, the following list of issues may be addressed: TFO / TrFO, Transcoding issues, DTX, Codec mode adaptation, Active Codec mode set handling, End-to-end delay handling.

7.5 Handling of inter-working between UTRAN and GERAN

There are a number of issues that differentiate UTRAN from GERAN when it comes to transport characteristics of real-time packet-switched data. Available bit-rate, service latency, packet loss patterns, and radio link performance, all contribute to IP-layer characteristics that are different from UTRAN. In order for multimedia telephony over IMS to be access-agnostic within 3GPP networks, it is important to take into account any difference that significantly influence service quality when inter-working between UTRAN and GERAN is taken into account.

7.6 Inter-media synchronization

Synchronization between speech and video is not a new problem for multimedia services and valid solutions exist. However, for real-time services such as Multimedia Telephony over IMS in 3GPP networks, it is possible that new situations arise that have not been addressed in standardization fora previously. Keeping real-time performance for all media components, while, at the same time accommodating for varying transport-channel characteristics for each separate stream, might induce synchronization problems that need to be carefully examined and, if needed, solved. Examples of such problems might be varying jitter between voice and video streams and loss of stream synchronization due to temporal media scaling, etc. QoS work performed elsewhere in 3GPP will be taken into account.

NOTE: The difference in packet sizes between voice and video may result in different jitter characteristics when transported over a shared channel, e.g. HSDPA.

7.7 SDU segmentation in UTRAN/GERAN

In UTRAN and GERAN networks IP packets are encapsulated in SDUs. These SDUs are mapped to PDUs that might serve as SDUs for the next underlying protocol layer. In UTRAN it is for example the case that IP packets are being encapsulated in PDCP-PDUs, which may become RLC-SDUs, which is mapped on RLC-PDUs. In GERAN similar segmentation takes place in corresponding layers. In general the PDU size is determined by some radio-bearer settings and is independent of the incoming SDU. However, whenever the SDU size exceeds the PDU size, the SDU needs to be segmented and reassembled at the corresponding layer at the receiver. The loss of a PDU generally results in the loss of all SDUs that are completely or partly included in the lost PDU. Typically, PDU sizes can be much smaller than SDU sizes, e.g. typical RLC-PDU sizes in GERAN and UTRAN range from 20 bytes to 80 bytes, which might result in multiple segments of an IP packet within the multimedia telephony service. In this case the SDU goodput might be significantly lower than the PDU goodput. By appropriate selection of SDU and/or PDU sizes the goodput might be optimized. Therefore, it is proposed to address and investigate the effects of segmentation in lower layers, which might lead to improvement in the goodput at the application layer.

7.8 Rate Adaptation

The Packet-switched Streaming Service (PSS) (3GPP TS 26.234 [28]) specifies rate adaptation mechanisms that enable streaming servers to make appropriate use of downlink resources given feedback signalled from a UE. Similar signalling capabilities may also be considered for conversational multimedia calls. When a particular UE is in a bad geometry, it may not only have poor media quality for itself, but also have an adverse effect on the latency of other users in the cell as well. Release-6 specifications of 3GPP TS 26.236 [4] specify RTCP for video calls but not for VoIP. Furthermore, default settings of RTCP may not be adequate for timely feedback of rate adaptation messages. It is proposed to investigate and provide appropriate rate adaptation signalling capabilities for multimedia telephony over IMS.

7.9 Packetization Overheads

The multimedia telephony service includes video, voice and possibly other media components for conversation in the future. Each media stream is delivered over RTP, requiring complete RTP/UDP/IP headers for each packet of every stream. Depending on the packetization schemes used, there is a trade-off between latency and the packetization overhead. Investigate and recommend appropriate guidelines for minimizing packetization overhead over the air.

7.10 End-2-End Signalling

Some solutions to problems addressed by the present document may require additional signalling. Investigate and specify appropriate signalling support.

8 Conclusions and recommendations

The present document has identified a number of areas of multimedia telephony over IMS where opportunities for optimization exist. The areas, identified in clause 7, include:

- delay jitter handling;

- packet-loss handling;

- handling of inter-working with CS;

- handling of inter-working between UTRAN and GERAN;

- inter-media synchronization;

- SDU segmentation in UTRAN/GERAN;

- rate adaptation;

- packetization overheads;

- end-2-end signalling.

The recommendations of the present document are:

- to perform work addressing the above areas;

- to create a new Technical Specification for an IMS Multimedia Telephony client in Release 7 supporting conversational speech, video and text with the scope to deliver a user experience comparable to that of CS conversational services using similar amount of resources.

Annex A:
Change history

<table>
<tr><td colspan="8" align="center">Change history</td></tr>
<tr><td>Date</td><td>TSG #SA</td><td>TSG Doc.</td><td>CR</td><td>Rev</td><td>Subject/Comment</td><td>Old</td><td>New</td></tr>
<tr><td>2006-03</td><td>31</td><td></td><td></td><td></td><td>TR 26.914 presented for approval</td><td>2.0.0</td><td>7.0.0</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
</table>

History

Document history		
V7.0.0	June 2007	Publication