I n t e r n a t i o n a l   T e l e c o m m u n i c a t i o n   U n i o n

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

J.388

(08/2010)

SERIES J: CABLE NETWORKS AND TRANSMISSION OF TELEVISION, SOUND PROGRAMME AND OTHER MULTIMEDIA SIGNALS

Digital transmission of television signals

Real-time video and audio transmission system over IP networks

Recommendation ITU-T J.388

ITU-T

Recommendation ITU-T J.388

Real-time video and audio transmission system over IP networks

Summary

Recommendation ITU-T J.388 defines a system for real-time video/audio transmission over an IP-based network. The protocol of the system consists of session control and media transmission. Session control is based on the session initiation protocol (SIP), and for the use of SIP this Recommendation refers to Recommendation ITU-T J.366.4. Several types of packet formats are defined for media transmission. With regard to codecs, this Recommendation provides some mandatory codecs and some optional ones. The system equipped with the protocol specified here can be used for bidirectional applications, such as video telephony, as well as unidirectional applications, such as video news gathering. Even in case of video news gathering, bidirectional transmission is often required for communication between a reporter in the field and the studio. In addition, the system is defined taking into account implementation complexity. Thus, a software-based embedded solution is also available, which requires less cost and a shorter development period than a hardware-based one.

History

Edition	Recommendation	Approval	Study Group
1.0	ITU-T J.388	2010-08-29	9

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications, information and communication technologies (ICTs). The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure, e.g., interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had not received notice of intellectual property, protected by patents, which may be required to implement this Recommendation. However, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database at http://www.itu.int/ITU-T/ipr/.

CONTENTS

Recommendation ITU-T J.388

Real-time video and audio transmission system over IP networks

1 Scope

This Recommendation defines a system for real-time video/audio transmission over an IP-based network. Three profiles are defined, for video telephony, SDTV video transmission, and HDTV video transmission. The protocol of the system consists of session control and media transmission. Session control is based on the session initiation protocol (SIP), and for the use of SIP this Recommendation refers to [ITU-T J.366.4]. Several types of packet formats are defined for media transmission. With regard to codecs, this Recommendation provides some mandatory codecs and some optional ones for each profile. The system equipped with the protocol specified here can be used for bidirectional applications such as video-enabled order wire as well as unidirectional applications such as electronic news gathering.

2 References

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published. The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

[ITU-T H.222.0] Recommendation ITU-T H.222.0 (2006) | ISO/IEC 13818-1:2007, *Information technology – Generic coding of moving pictures and associated audio information: Systems*.

[ITU-T H.262] Recommendation ITU-T H.262 (2000) | ISO/IEC 13818-2:2000, *Information technology – Generic coding of moving pictures and associated audio information: Video*.

[ITU-T H.263] Recommendation ITU-T H.263 (2005), *Video coding for low bit rate communication*.

[ITU-T H.264] Recommendation ITU-T H.264 (2010), *Advanced video coding for generic audiovisual services*.

[ITU-T J.360] Recommendation ITU-T J.360 (2006), *IPCablecom2 architecture framework.*

[ITU-T J.361] Recommendation ITU-T J.361 (2006), *IPCablecom 2 codec and media*.

[ITU-T J.366.3] Recommendation ITU-T J.366.3 (2010), *IPCablecom2 IP Multimedia Subsystem (IMS); Stage 2 specification*.

[ITU-T J.366.4] Recommendation ITU-T J.366.4 (2010), *IPCablecom2 IP Multimedia Subsystem (IMS): Session Initiation Protocol (SIP) and Session Description Protocol (SDP) – Stage 3 Specification*.

[ITU-T J.366.7] Recommendation ITU-T J.366.7 (2010), *IPCablecom2 IP Multimedia Subsystem (IMS): Access security for IP-based services*.

[ETSI EN 301 704] ETSI EN 301 704 V.7.2.1 (2000), *Digital cellular telecommunications system (Phase 2+) (GSM); Adaptive Multi-Rate (AMR) speech transcoding*.

[IETF RFC 2250]	IETF RFC 2250 (1998), *RTP Payload Format for MPEG1/MPEG2 Video*.
[IETF RFC 2327]	IETF RFC 2327 (1998), *SDP: Session Description Protocol*.
[IETF RFC 3016]	IETF RFC 3016 (2000), *RTP Payload Format for MPEG-4 Audio/Visual Streams*.
[IETF RFC 3261]	IETF RFC 3261 (2002), *SIP: Session Initiation Protocol*.
[IETF RFC 3267]	IETF RFC 3267 (2002), *Real-Time Transport Protocol (RTP) Payload Format and File Storage Format for the Adaptative Multi-Rate (AMR) and Adaptative Multi-Rate Wideband (AMR-WB) Audio Codecs*.
[IETF RFC 3550]	IETF RFC 3550 (2003), *RTP: A Transport Protocol for Real-Time Applications.*
[IETF RFC 3984]	IETF RFC 3984 (2005), *RTP Payload Format for H.264 Video*.
[IETF RFC 4629]	IETF RFC 4629 (2007), *RTP Payload Format for ITU-T Rec. H.263 Video*.
[ISO/IEC 13818-3]	ISO/IEC 13818-3:1998, *Information technology – Generic coding of moving pictures and associated audio information – Part 3: Audio*.
[ISO/IEC 13818-7]	ISO/IEC 13818-7:2006, *Information technology – Generic coding of moving pictures and associated audio information – Part 7: Advanced Audio Coding (AAC)*.
[ISO/IEC 14496-2]	ISO/IEC 14496-2:2004, *Information technology – Coding of audio-visual objects – Part 2: Visual*.
[ISO/IEC 14496-3]	ISO/IEC 14496-3:2009, *Information technology – Coding of audio-visual objects – Part 3: Audio*.

3 Definitions

3.1 Terms defined elsewhere

This Recommendation uses the following terms defined elsewhere:

3.1.1 message [IETF RFC 3261]: Data sent between SIP elements as part of the protocol. SIP messages are either requests or responses.

3.1.2 method [IETF RFC 3261]: The method is the primary function that a request is meant to invoke on a server. The method is carried in the request message itself. Example methods are INVITE and BYE.

3.1.3 proxy, proxy server [IETF RFC 3261]: An intermediary entity that acts as both a server and a client for the purpose of making requests on behalf of other clients. A proxy server primarily plays the role of routing, which means its job is to ensure that a request is sent to another entity "closer" to the targeted user. Proxies are also useful for enforcing policy (for example, making sure a user is allowed to make a call). A proxy interprets, and, if necessary, rewrites specific parts of a request message before forwarding it.

3.1.4 request [IETF RFC 3261]: A SIP message sent from a client to a server, for the purpose of invoking a particular operation.

3.1.5 response [IETF RFC 3261]: A SIP message sent from a server to a client, for indicating the status of a request sent from the client to the server.

3.1.6 server [IETF RFC 3261]: A server is a network element that receives requests in order to service them and sends back responses to those requests. Examples of servers are proxies, user agent servers, redirect servers, and registrars.

3.1.7 session [IETF RFC 3261]: From the SDP specification: "A multimedia session is a set of multimedia senders and receivers and the data streams flowing from senders to receivers. A multimedia conference is an example of a multimedia session". [IETF RFC 2327] (A session as defined for SDP can comprise one or more RTP sessions.) As defined, a callee can be invited several times, by different calls, to the same session. If SDP is used, a session is defined by the concatenation of the SDP user name, session id, network type, address type, and address elements in the origin field.

3.2 Terms defined in this Recommendation

This Recommendation defines the following term:

3.2.1 profile: A specified subset of the functionalities of the terminal defined in this Recommendation, or a specified subset of the bitstream syntax of video codecs.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

AAC Advanced Audio Coding

AMR Adaptive Multi-Rate

AV Audio/Video

AVC Advanced Video Coding

BP Baseline Profile

CSCF Call Session Control Function

ENG Electronic News Gathering

HDTV High Definition Television

HL High Level

HP High Profile

I-CSCF Interrogating-CSCF

IMS IP Multimedia Subsystem

IP Internet Protocol

L* Level* (e.g., L0 means Level Zero)

MAC Medium Access Control

ML Main Level

MP Main Profile

NB Narrow-Band

P0 Profile Zero

P-CSCF Proxy-CSCF

RTP Real-time Transport Protocol

SBR Spectral Band Replication

S-CSCF Serving-CSCF

SDP Session Description Protocol

SDTV Standard Definition Television

SIP	Session Initiation Protocol
SP	Simple Profile
TCP	Transmission Control Protocol
TS	Transport Stream
UDP	User Datagram Protocol
UE	User Equipment
URI	Uniform Resource Identifier
VT	Video Telephony (bidirectional conversational communication with video and audio)

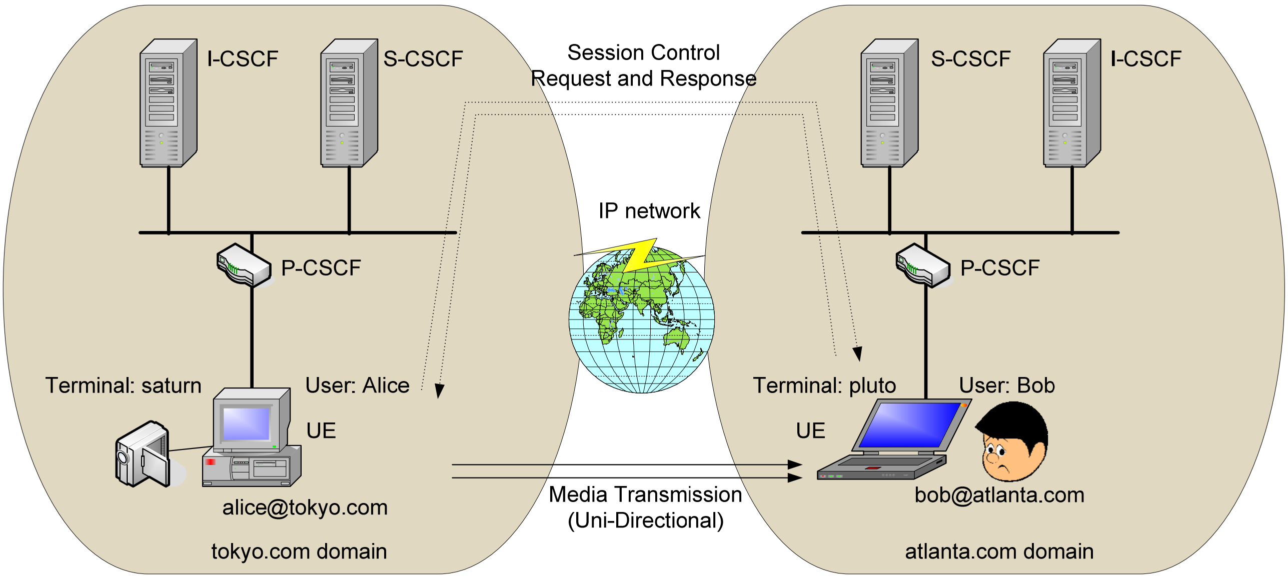
5 Conventions

None.

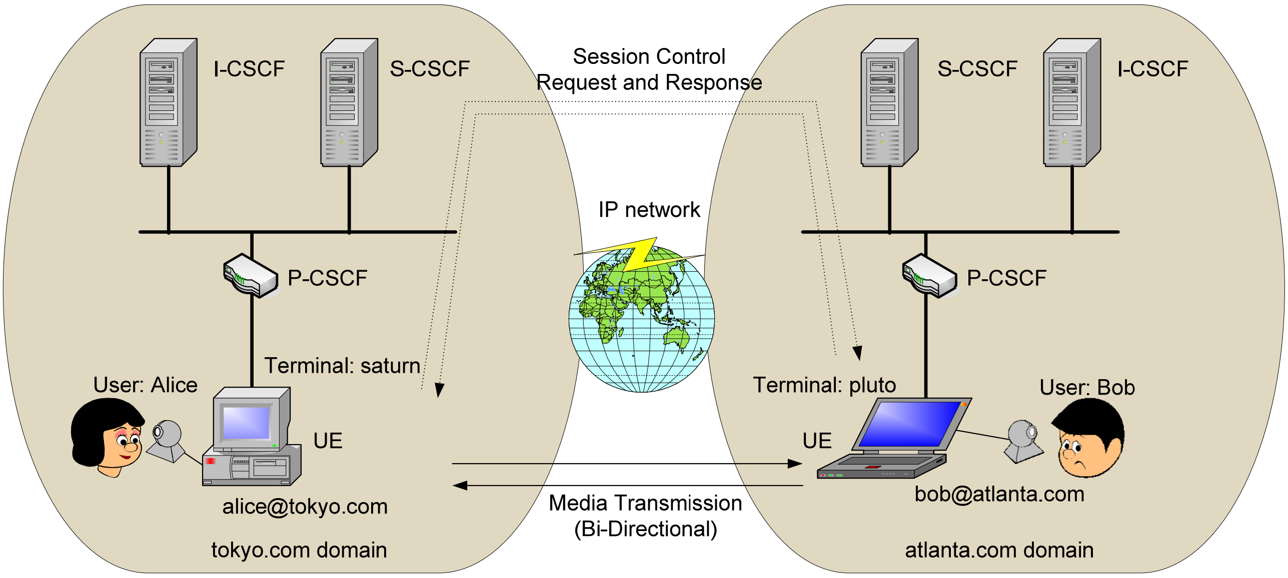
6 System overview

6.1 Architecture framework

This Recommendation specifies protocols for session control and media transmission for real-time video/audio transmission over IP networks. Examples of the system architecture are shown in Figure 1, and an example protocol stack is shown in Figure 2.



a) Unidirectional case



b) Bidirectional case

Figure 1 – Examples of system architecture

Session control (SIP)	Codec
	Packet format
UDP or TCP	UDP
IP	
MAC	
Physical	

Figure 2 – Example protocol stack of the system

As shown in these examples, protocols defined in this Recommendation can be used for unidirectional applications such as video news gathering and bidirectional applications such as video telephony.

The system contains the following components:

- Terminal (UE): A terminal issues SIP requests and receives SIP responses for session set-up, and transmits and receives media packets of video and audio.
- Registrar: A registrar provides address binding for a particular domain. S-CSCF may behave as a registrar as defined in [ITU-T J.360].
- Proxy: A proxy routes SIP messages from the UE to the I-CSCF or S-CSCF and vice versa. This is performed by the P-CSCF.

Detailed descriptions of the components are given by [ITU-T J.360]. This Recommendation focuses on the specification of terminals. The use of registrar and proxy is optional. A terminal defined in this Recommendation shall support communications without these servers, in other words, peer-to-peer communication.

6.2 Session control

The session initiation protocol (SIP) is used for creating, modifying and terminating sessions with users in this Recommendation.

This Recommendation uses the following SIP methods, extracted from the methods defined in [ITU-T J.366.4]:

- REGISTER
- INVITE
- ACK
- BYE
- CANCEL
- PRACK
- INFO

A terminal conforming to this Recommendation shall have the capability to send these requests and receive responses directly peer to peer without a proxy, as well as via a proxy.

6.3 Media transmission

After a session has been established, the terminals start sending media packets using the format agreed in the SDP, which is exchanged in the INVITE request. In general, the end-to-end media packets are directly transmitted and take a path different from that of the SIP signalling messages.

In this Recommendation, a terminal shall support both unidirectional transmission and bidirectional transmission.

Clause 8.2 defines the types of codecs used in this Recommendation, and clause 8.4 defines the packet formats used in this Recommendation.

6.4 Example of the sequence

The following is a typical example of a SIP message exchange between two users, Alice and Bob. In this example, it is assumed that Alice and Bob have already completed the registration.

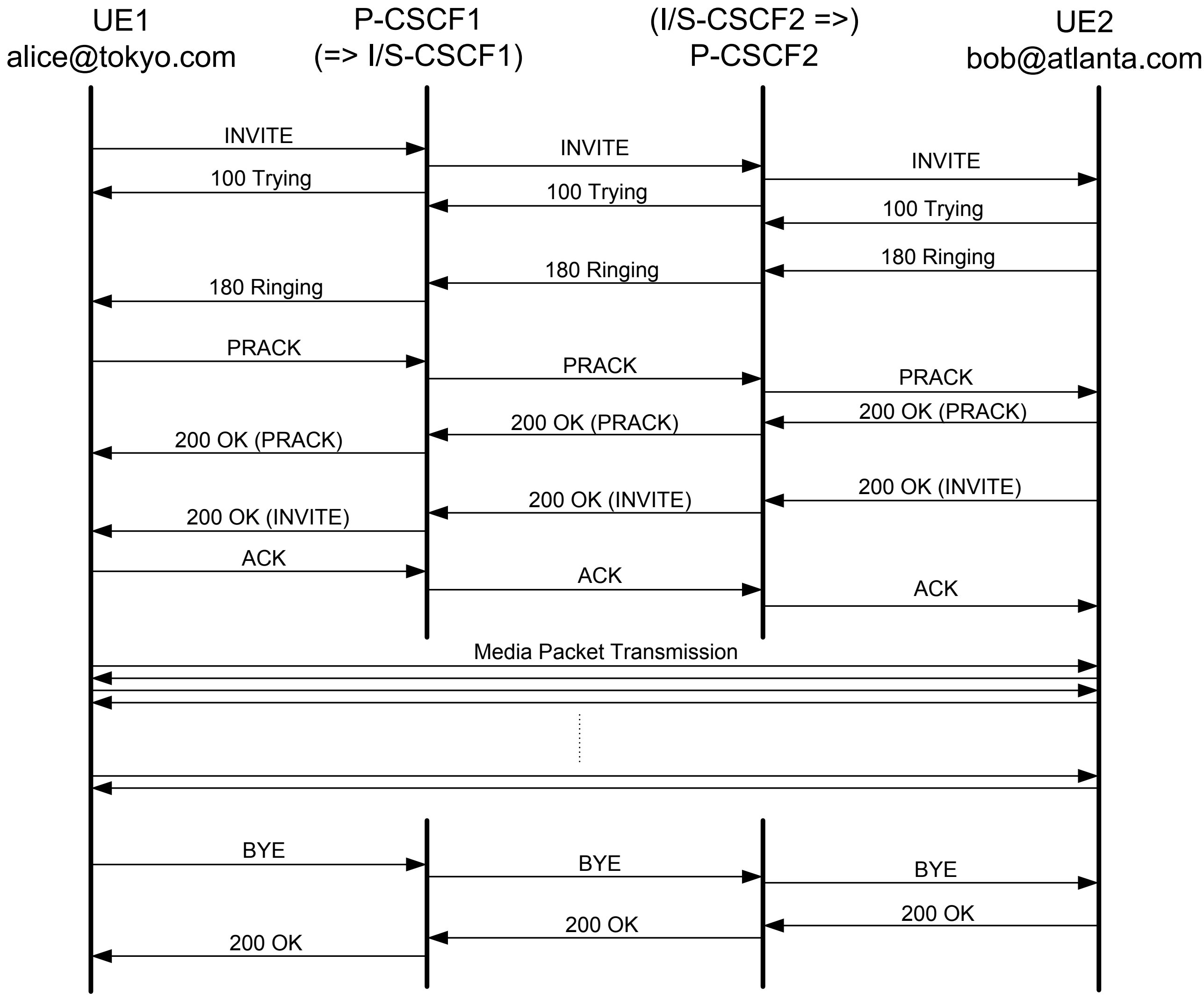


Figure 3 – Example sequence

First of all, Alice calls Bob using his SIP identity, which is called a SIP URI. The SIP URI used in this Recommendation has a form similar to an e-mail address, containing a user name and a host name. "sip:" is used as a scheme name. In this case, SIP URI for Alice is "sip:alice@tokyo.com", and that for Bob is "sip:bob@atlanta.com," where "tokyo.com" and "atlanta.com" are the domains of Alice's and Bob's SIP service provider, respectively. In this example, the transaction begins with an INVITE request addressed to Bob's SIP URI from Alice.

Detailed sequences are described in [ITU-T J.366.4], where the message formats are also specified.

7 Session control

This clause explains the outline of SIP functionalities. Details of SIP session control are specified in [ITU-T J.366.4].

7.1 Authentication

A terminal, a proxy server, and a registrar shall support the SIP digest authentication mechanism defined in [ITU-T J.366.7].

7.2 Registration

A terminal and a registrar shall support the registration mechanism defined here. In the example shown in Figure 1, a registrar is separated from a proxy server. The functionality of a registrar, however, may be collocated in a proxy server. In this case, a terminal sends a REGISTER request to the same server as that to which it sends an INVITE request.

7.3 Starting a session

An INVITE request is used to begin a session. Terminals and proxy servers shall support the session starting mechanism defined here.

7.4 Closing a session

A BYE request is used to tear down a session already established. Terminals and proxy servers shall support the session closing mechanism defined here.

7.5 Cancelling a request

A CANCEL request is used to cancel a SIP request already sent but not answered. A terminal and a proxy server shall support the cancel request mechanism defined here.

8 Media transmission

8.1 Profiles

This Recommendation defines the following three profiles.

- VT Profile: Mainly used for video telephony applications.
- SD Profile: Mainly used for SDTV ENG applications.
- HD Profile: Mainly used for HDTV ENG applications.

Table 1 defines the combination of codecs and packet types according to the definitions in [ITU-T J.361].

Table 1 – Combination of codecs and packet types

Profile	Video	Audio	Transport
VT	M: MPEG-4 SP@L0 O: H.264 BP@L1.2 O: H.263 P0@L10	M: AMR-NB	M: RTP M: AV Integrated
SD	M: H.264 MP@L3	M: MPEG-2 AAC O: MPEG-4 AAC SBR	M: RTP (TS over RTP)
	M: H.262 MP@ML	M: MPEG-2 AAC	M: RTP (TS over RTP)
HD	M: H.264 HP@L4.0	M: MPEG-2 AAC O: MPEG-4 AAC SBR	M: RTP (TS over RTP)
	M: H.262 MP@HL	M: MPEG-2 AAC	M: RTP (TS over RTP)
M – Mandatory O – Optional			

Terminals may support real-time text conversation with video and audio media as defined in clause 8.6.

8.2 Codecs

In this clause, the types of codecs and their profiles and levels used in this Recommendation are defined.

8.2.1 ITU-T H.264

MPEG-4 AVC [ITU-T H.264] Baseline Profile at Level 1.2 codec is optional for VT Profile. Main Profile at Level 3 for SD Profile and High Profile at Level 4.0 for HD Profile are mandatory in this Recommendation.

8.2.2 ITU-T H.262 (MPEG-2 Video)

MPEG-2 Video [ITU-T H.262] is mandatory for SD Profile and HD Profile in this Recommendation. If MPEG-2 video is used, audio shall be MPEG-2 AAC [ISO/IEC 13838-7]. In this case, both elementary streams shall be multiplexed in MPEG-2 TS format [ITU-T H.222.0].

8.2.3 MPEG-4 Visual

MPEG-4 Visual [ISO/IEC 14496-2] Simple Profile at Level 0 codec is mandatory for all terminals that are capable of VT Profile. Other profiles and levels of MPEG-4 visual are optional, which may be exchanged in a SDP message of an INVITE request.

8.2.4 ITU-T H.263

[ITU-T H.263] Profile 0 Level 10 codec is optional for VT Profile in this Recommendation.

8.2.5 MPEG-4 Audio

MPEG-4 AAC-SBR and other types of MPEG-4 Audio [ISO/IEC 14496-3] are optional for SD and HD Profiles, which may be exchanged in a SDP message of an INVITE request.

8.2.6 AMR-NB

ETSI AMR Narrow Band speech codec [ETSI EN 301 704] is mandatory for all terminals that are capable of VT Profile.

8.3 Codec profile selection in set-up – UE procedures

The network might reject a session set-up issued by a UE due to lack of network resources corresponding to the required codec profile. In this case, the UE shall fail to complete the session with that codec profile and should attempt to complete the session with a different profile.

A UE may also terminate the session and start over with a different profile according to the transmission status such as packet jitters or packet losses monitored by the UE. This is implementation dependent and outside the scope of this Recommendation.

8.4 Packet format

In this clause, types of packet format that convey media bitstream over an IP network are defined. The type of the packet format shall be indicated in the SDP message. A terminal is required to handle uneven packet transmissions, where the packet format for the sending direction is different from that of the receiving direction.

8.4.1 RTP

A terminal shall support an RTP packet format defined in [IETF RFC 3550].

Payload formats for RTP shall be conformant with the following specifications.

MPEG-4 Part 2: [IETF RFC 3016] shall be used for VT Profile. The bitstream shall be conveyed in the TS packets for the SD and HD profiles.

ITU-T H.264 | MPEG-4 Part 10: [IETF RFC 3984] shall be used for VT Profile.

ITU-T H.263: [IETF RFC 4629] Type A shall be used for VT Profile.

AMR-NB: [IETF RFC 3267] shall be used for VT Profile.

MPEG-TS: [IETF RFC 2250] shall be used for SD and HD Profiles.

MPEG-2 Video: Bitstream shall be conveyed in the TS packets [ITU-T H.222.0].

MPEG-2 AAC Audio: Bitstream shall be conveyed in the TS packets [ITU-T H.222.0].

MPEG-4 AAC-SBR Audio: Bitstream shall be conveyed in the TS packets [ITU-T H.222.0].

8.4.2 A/V integrated format (non-RTP)

This Recommendation defines a packet format where audio data and video data are multiplexed in one packet. This format provides less packet overhead and simple AV synchronization at a decoder. A terminal for VT Profile shall support the A/V integrated packet format defined here. The packet shall contain an integer number of audio encoded frames. The size of the video bitstream is strongly recommended to correspond to the equivalent time of the audio frames. For example, if the audio frames are equivalent to 60 ms (3 AMR-NB frames), the size of video should approximately be BR $\times$ 0.06.

IP header 20 bytes	UDP header 8 bytes	Serial No. 1 byte	Audio data size 1 byte	Audio bitstream	Video bitstream

8.5 Transport

Media packets are transmitted over UDP/IP. UDP/IP is also recommended for session control packets.

8.6 Real-time text support (optional)

The protocol for multimedia application text conversation defined in [ITU-T J.361] may be optionally supported. It introduces real-time conversational text as a type of media in multimedia communication. If this option is used, the type of transport of real-time text shall be RTP and its payload format defined in [ITU-T J.361] shall be used.

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
Series D	General tariff principles
Series E	Overall network operation, telephone service, service operation and human factors
Series F	Non-telephone telecommunication services
Series G	Transmission systems and media, digital systems and networks
Series H	Audiovisual and multimedia systems
Series I	Integrated services digital network
Series J	**Cable networks and transmission of television, sound programme and other multimedia signals**
Series K	Protection against interference
Series L	Construction, installation and protection of cables and other elements of outside plant
Series M	Telecommunication management, including TMN and network maintenance
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	Terminals and subjective and objective assessment methods
Series Q	Switching and signalling
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminals for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	Data networks, open system communications and security
Series Y	Global information infrastructure, Internet protocol aspects and next-generation networks
Series Z	Languages and general software aspects for telecommunication systems

Printed in Switzerland
Geneva, 2010