



INTERNATIONAL TELECOMMUNICATION UNION

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

J.120

(05/2000)

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

Interactive systems for digital television distribution

**Distribution of sound and television programs
over the IP network**

ITU-T Recommendation J.120

(Formerly CCITT Recommendation)

ITU-T J-SERIES RECOMMENDATIONS

TRANSMISSION OF TELEVISION, SOUND PROGRAMME AND OTHER MULTIMEDIA SIGNALS

General Recommendations	J.1–J.9
General specifications for analogue sound-programme transmission	J.10–J.19
Performance characteristics of analogue sound-programme circuits	J.20–J.29
Equipment and lines used for analogue sound-programme circuits	J.30–J.39
Digital encoders for analogue sound-programme signals	J.40–J.49
Digital transmission of sound-programme signals	J.50–J.59
Circuits for analogue television transmission	J.60–J.69
Analogue television transmission over metallic lines and interconnection with radio-relay links	J.70–J.79
Digital transmission of television signals	J.80–J.89
Ancillary digital services for television transmission	J.90–J.99
Operational requirements and methods for television transmission	J.100–J.109
Interactive systems for digital television distribution	J.110–J.129
Transport of MPEG-2 signals on packetised networks	J.130–J.139
Measurement of the quality of service	J.140–J.149
Digital television distribution through local subscriber networks	J.150–J.159

For further details, please refer to the list of ITU-T Recommendations.

Distribution of sound and television programs over the IP network

Summary

This Recommendation defines the transmission protocol and system configuration for distributing sound and television programs over the Internet "Webcasting". It specifies the operations necessary to adapt audio and video bitstreams to the Internet protocol and the functional characteristics associated with this system.

This Recommendation includes an electronic attachment containing sample source code and some tools for conformance tests.

Source

ITU-T Recommendation J.120 was prepared by ITU-T Study Group 9 (1997-2000) and approved under the WTSC Resolution 1 procedure on 18 May 2000.

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications. 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 Conference (WTSC), 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 WTSC 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.

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, implementors are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database.

© ITU 2001

All rights reserved. No part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from ITU.

CONTENTS

	Page
1 Scope.....	1
2 References.....	1
2.1 Normative references	1
2.2 Informative references	1
3 Terms and definitions	2
4 Abbreviations.....	2
5 System definition	3
5.1 Procedures for the receiver	4
5.1.1 Acquisition of a presentation description	4
5.1.2 Downloading audio/video decoders	4
5.1.3 Setting-up a server	4
5.1.4 Receiving streams.....	4
5.2 Video decoding	4
5.3 Audio decoding	4
5.4 Application Programming Interface (API) of the decoder.....	5
5.5 Network interface.....	5
6 Session control.....	6
6.1 Overall operation of the session control	6
6.2 Presentation description	7
6.2.1 "v=" field	7
6.2.2 "o=" field	7
6.2.3 "s=" field.....	7
6.2.4 "i=" field	7
6.2.5 "m=" field	8
6.2.6 "a=" field.....	8
6.3 RTSP control methods	8
6.3.1 SETUP	8
6.3.2 PLAY	8
6.3.3 PAUSE.....	9
6.3.4 TEARDOWN	9
6.4 RTSP header fields	9
6.4.1 CSeq.....	9
6.4.2 Range	9
6.4.3 Session.....	10
6.4.4 Transport.....	10

	Page
7 RTP header formats	10
Appendix I – Simple example of RTSP presentation description and session control.	12
Appendix II – A typical RTP payload format for ITU-T H.263	13
Appendix III – An example of RTP payload format for audio data.....	15
Appendix IV – Preferred embodiments of downloading decoders	16
Appendix V – Sample Software.....	17
Electronic attachment: source code and conformance test tools	

ITU-T Recommendation J.120

Distribution of sound and television programs over the IP network¹

1 Scope

This Recommendation defines the transmission protocol and system configuration for distributing sound and television programs over the Internet "Webcasting". This concerns the end-to-end signal flow from the program server to user receiver. This transmission chain contains the signal encoding/decoding, packet mapping, session control and network transmission.

The following items are specified:

- Presentation description which provides the overall program and the properties of the media.
- Server control protocol which provides signalling for proper operation between the server and receiver.
- Stream packet mapping which formats coded video and audio streams into messages for output to the network interface and retrieves received video and audio data.
- Required receiver capability.
- Downloadable software decoders.

This Recommendation is restricted to the Internet Protocol (IP) as a network layer. Usage of transport layer protocol is outside its scope.

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; all 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.

2.1 Normative references

- ITU-T H.225.0 (2000), *Call signalling protocols and media stream packetization for packet-based multimedia communication systems*.

2.2 Informative references

- ITU-T G.723.1 (1996), *Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s*.
- ITU-T H.222.0 (2000) | ISO/IEC 13818-1:1996, *Information technology – Generic coding of moving pictures and associated audio information: Systems*.
- ITU-T H.262 (2000) | ISO/IEC 13818-2:1996, *Information technology – Generic coding of moving pictures and associated audio information: Video*.
- ITU-T H.263 (1998), *Video coding for low bit rate communication*.

¹ This Recommendation includes an electronic attachment containing sample source code and some tools for conformance tests.

- ISO/IEC 13818-3:1998, *Information technology – Generic coding of moving pictures and associated audio information – Part 3: Audio*.
- ISO/IEC 14496-2:1999, *Information technology – Coding of audio-visual objects – Part 2: Visual*.
- ISO/IEC 14496-3:1999, *Information technology – Coding of audio-visual objects – Part 3: Audio*.
- RFC 2190 (1997), *RTP Payload Format for H.263 Video Streams*.
- RFC 2198 (1997), *RTP Payload for Redundant Audio Data*.
- RFC 2326 (1998), *(RTSP), Real Time Streaming Protocol (RTSP)*.

3 Terms and definitions

This Recommendation defines the following terms:

- 3.1 Internet Protocol (IP):** An Internet network-layer protocol, defined by the IETF.
- 3.2 IP address:** The network layer address defined by the Internet Protocol. This address is mapped onto the layer one address of the respective system.
- 3.3 Port:** The abstraction that transport protocols use to distinguish among multiple destinations within a given host computer. The transport selectors used by the OSI transport layers are equivalent to ports.
- 3.4 Real-time Transport Protocol (RTP):** A transport protocol for real-time applications defined in ITU-T H.225.0. It is designed for real-time transmission of audio and video data.
- 3.5 Real-time Streaming Protocol (RTSP):** RTSP is a protocol defined in 6.3 and 6.4. RTSP specifies session control method between server and receiver. Its syntax is based on HTTP.
- 3.6 Transport Stream (TS):** A data packet possessing a length of 188 bytes including 4 bytes of header information. The header contains MPEG related data.
- 3.7 Webcasting:** Distribution of sound and television programs over the Internet.

4 Abbreviations

This Recommendation uses the following abbreviations:

API	Application Programming Interface
FTP	File Transport Protocol
GOB	Group of Blocks
HTTP	Hyper Text Transport Protocol
IETF	Internet Engineering Task Force
IP	Internet Protocol
IPv4	Internet Protocol Version 4
IPv6	Internet Protocol Version 6
MPEG	Moving Picture Experts Group
NPT	Normal Play Time
RTCP	Real-time Transport Control Protocol
RTP	Real-time Transport Protocol

RTSP	Real-time Streaming Protocol
SDP	Session Description Protocol
TCP	Transmission Control Protocol
TS	Transport Stream
UDP	User Datagram Protocol
URI	Universal Resource Identifier
URL	Uniform Resource Locators

5 System definition

The system and its signal flow in Webcasting are shown in Figure 1.

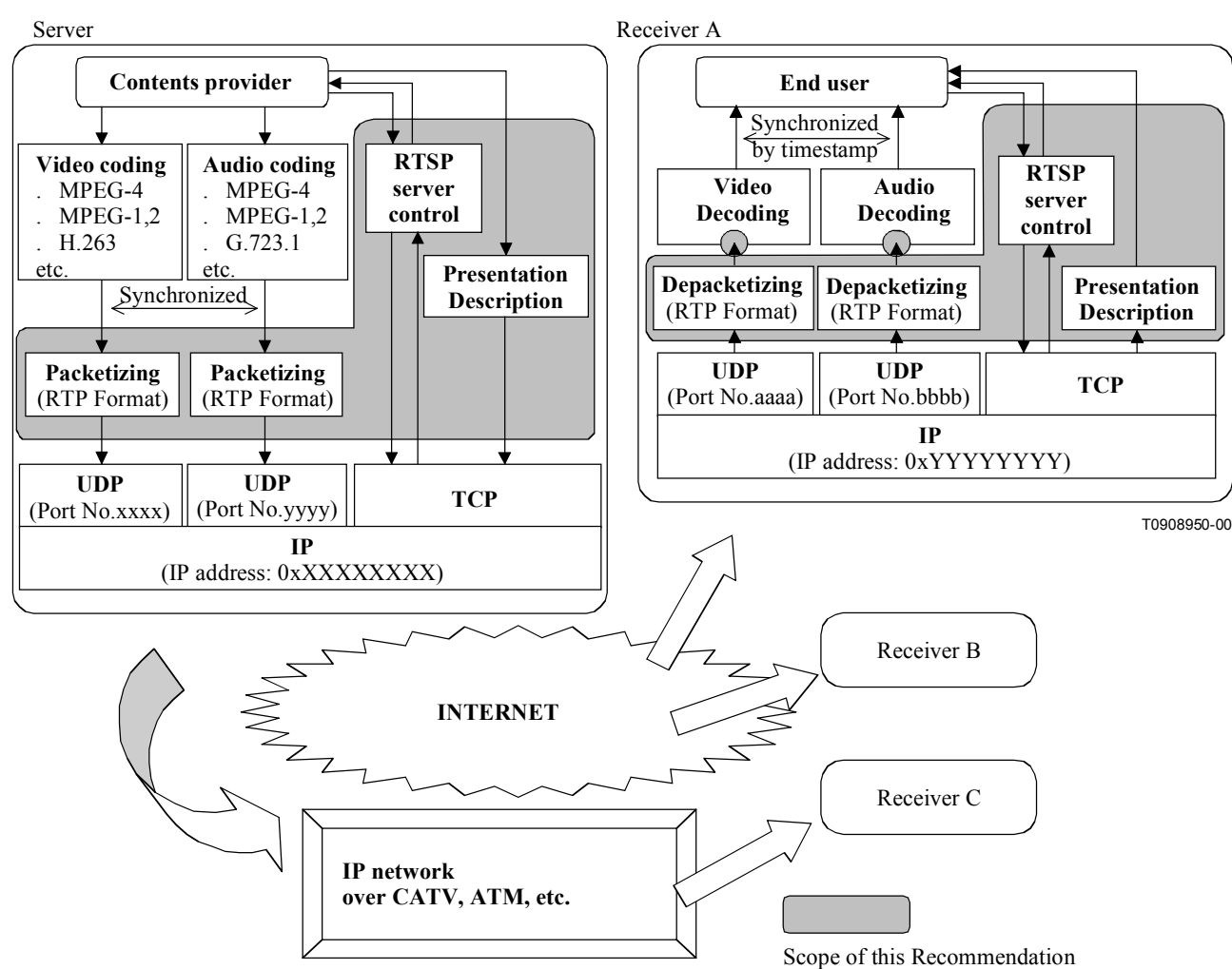


Figure 1/J.120 – Block diagram of Webcasting system

5.1 Procedures for the receiver

5.1.1 Acquisition of a presentation description

The presentation description defines the overall program and the properties of the media in the form of the description defined in 6.2. The acquisition method of the presentation description may be HTTP, FTP or others such as e-mail, but a definition is outside the scope of this Recommendation.

The receiver may obtain information relating to program contents, details of streams constituting the program, decoder downloading information and other parameters.

5.1.2 Downloading audio/video decoders

After acquisition of the presentation description, if necessary the receiver may obtain the audio/video decoder by downloading it from the server defined in the presentation description or from storage media such as a CD-ROM.

Downloading information is specified in the <a> field of the presentation description, it may directly indicate the decoder software by HTTP or FTP or it may be linked to the HTML script of the downloading page. The receiver which is capable of downloading decoders shall be capable of accessing the HTML 4.0 script and of downloading decoder software by HTTP and FTP.

5.1.3 Setting-up a server

After decoding the presentation description as well as after downloading the decoder(s) if necessary, the receiver may send audio/video stream controls in the form of RTSP to the stream servers specified in the presentation description. The receiver shall be equipped with the minimal RTSP implementation.

In the setting-up process, RTSP may not deliver the audio/video streams itself, but instead act as a "network remote control" for stream servers.

The SETUP control method of RTSP may start an RTSP session and may cause the server to prepare to send the stream. The PLAY method may start data transmission on a stream specified via SETUP. The TEARDOWN method frees resources associated with the stream.

5.1.4 Receiving streams

Actual audio/video streams shall be sent via separate protocol RTP, independent of the session control channel. Properties of transport protocol such as the port number may be obtained from the RTSP session control prior to receiving streams. The receiver shall be equipped with RTP/RTCP capability. Received packets from each stream are transferred to the decoders. The application programming interface (API) for the audio/video decoders is specified in clause 5.4.

5.2 Video decoding

The video coding method and its downloadable sites are identified in the presentation description. More than one video channel may be transmitted, as negotiated via the RTSP session set-up. The video stream packet is formatted as described in clause 7.

The video stream packet is transferred to the video decoder via the application programming interface described in clause 5.4. All video decoders shall have the application programming interface.

5.3 Audio decoding

The audio coding algorithm and its downloadable sites shall be specified in the presentation description. The receivers may optionally receive more than one audio channel at the same time, for example, to allow two languages or stereo sound to be conveyed.

Audio packets should be delivered to the transport layer periodically at an interval determined by the audio codec Recommendation in use (audio frame interval). The delivery of each audio packet should occur no later than 5 ms after a whole multiple of the audio frame interval, measured from delivery of the first audio frame (audio delay jitter). The audio packets are formatted as described in clause 7.

The audio stream packet is transferred to the video decoder via application programming interface described in clause 5.4. All audio decoders shall have the application programming interface.

5.4 Application Programming Interface (API) of the decoder

A decoder program shall have functions which transfer audio/video packets to the decoder. The interface functions are specified as below.

```
extern "C" {  
    int putPacketStream(unsigned char *buffer, int length);  
}
```

Description:

- Transferring a packet stream into the decoder program.

Parameters:

- unsigned char *buffer: address to a packet stream
- int length: length of the packet

Return value:

- 0: succeeded
- -1: failed (input buffer over flow)

```
extern "C" {  
    int getBufferFreeSize(void)  
}
```

Description:

- This function returns free space in the internal packet buffer of the decoder.

Return value:

- Free space in bytes.

5.5 Network interface

The network interface is an implementation specific issue and is outside of the scope of this Recommendation. However, the network interface shall provide services including the following: Reliable (e.g. TCP) end-to-end service is mandatory for the RTSP control channel. Unreliable (e.g. UDP) end-to-end service is mandatory for the audio channels and the video channels. These services may be duplex or simplex, unicast or multicast depending on the capabilities of the terminals and the configuration of the network.

6 Session control

6.1 Overall operation of the session control

The overall program and the properties of the media shall be defined by a presentation description file. The presentation description file may be obtained by the receiver using HTTP or other means such as e-mail and may not necessarily be stored on the server.

The presentation description file contains a description of the media streams making up the program, including their encoding types, language, and other parameters that enables the receiver to choose the most appropriate media combination.

Fields used in the presentation description are summarized in Table 1.

Table 1/J.120 – The presentation description

Field name	Description
v	The version of the Session Description Protocol.
o	Information on the originator of the session. It gives their username, the address of the user's host, session ID and session version.
s	The session name.
i	Informative description of the session.
m	The media description. Media type, port number, transport protocol and payload types.
a	Attributes of the session or the media. It may specify the control URL, decoder download sites' URL or language of the session or the media.

Details of the fields are described in 6.2.

After decoding the presentation description, the RTSP controls media streams which may be sent via a separate protocol, independent of the control channel. For example, RTSP control may occur on a TCP connection while the data flows via UDP. Thus, data delivery continues even if no RTSP requests are received by the media server. Also, during its lifetime, a single media stream may be controlled by RTSP requests issued sequentially on different TCP connections. The control methods are listed in Table 2.

Table 2/J.120 – RTSP control methods

Methods	Description
SETUP	Causes the server to allocate resources for a stream and start an RTSP session.
PLAY	Starts data transmission on a stream allocated via SETUP.
PAUSE	Temporarily halts a stream without freeing server resources.
TEARDOWN	Frees resources associated with the stream.

Details of these methods are described in 6.3.

These methods are issued with header fields. The headers are summarized in Table 3.

Table 3/J.120 – RTSP header fields

Headers	Description
CSeq	The sequence number for an RTSP request-response pair.
Range	Range of time for a PLAY and a PAUSE.
Session	An identifier of the session started by SETUP and concluded by TEARDOWN.
Transport	Protocol to be used.

6.2 Presentation description

The definition and syntax are described below.

6.2.1 "v=" field

The "v=" field gives the version of the Session Description Protocol. There is no minor version number.

6.2.2 "o=" field

The "o=" field gives the originator of the session (their username and the address of the user's host) plus a session ID and session version number.

`o=<username> <session ID> <version> <network type> <address type> <address>`

`<username>` is the user's login on the originating host, or it is "-" if the originating host does not support the concept of a user ID. `<username>` does not contain spaces.

`<session ID>` is a numeric string globally unique to the session. The method of `<session ID>` allocation is suggested in which a Network Time Protocol (NTP) timestamp be used to ensure uniqueness.

`<version>` is the version number for this announcement. It is needed for proxy announcements to detect which of several announcements for the same session is the most recent. `<version>` is increased when a modification is made to the session data. It is also recommended (but not mandatory) that an NTP timestamp be used.

`<network type>` is a text string giving the type of network. Initially "IN" is defined to have the meaning "Internet".

`<address type>` is a text string giving the type of the address that follows. Initially "IP4" and "IP6" are defined.

`<address>` is the globally unique address of the machine from which the session was created. For an IPv4 address type, this is either a fully-qualified domain name of the machine or the dot-decimal representation of the IP version 4 address of the machine. For an IPv6 address type, this is either the fully-qualified domain name of the machine or the compressed textual representation of the IP version 6 address of the machine.

6.2.3 "s=" field

The "s=" field is the session name. There shall be one and only one "s=" field per session description.

6.2.4 "i=" field

The "i=" field is information about the session. There may be at most one session-level "i=" field per session description, and at most one "i=" field per media.

6.2.5 "m=" field

The "m=" field is the media announcement. Each media description starts with an "m=" field, and is terminated by either the next "m=" field or by the end of the session description.

m=<media> <port> <transport> <payload type(s)>

<media> is the media type. "audio" and "video" are defined.

<port> is the port number which serves as a recommendation from the server to the receiver. The receiver has to include this in its SETUP request and may ignore this recommendation. If the server has no preference, it should set the port number value to zero.

<transport> is the transport protocol. It is always set to "RTP/AVP" which means the Real-time Transport Protocol using the Audio/Video profile carried over UDP.

<payload type(s)> and subsequent fields are the payload type(s) of the media. The definition is listed in ITU-T H.225.0.

6.2.6 "a=" field

The "a=" field specifies attributes of the session or the media.

<a=control:> is the attribute used to convey the control URL. This attribute is used both for the session and media descriptions. If used for individual media, it indicates the URL to be used for controlling that particular media stream. If found at the session level, the attribute indicates the URL for aggregate control. This attribute may contain either relative or absolute URLs.

<a=decoder:> is the attribute used to convey the URL of decoder download sites. It indicates the URL to be accessed for downloading decoder software which a receiver does not have. This attribute may directly indicate the decoder software by HTTP or FTP, or it may be linked to the HTML script of the downloading page. The receiver which has the capability to download decoders shall be capable of accessing the HTML 4.0 script and of downloading decoder software by HTTP and FTP.

6.3 RTSP control methods

The RTSP control methods indicate the method to be performed on the resource identified by the Request-URI. The method is case-sensitive.

6.3.1 SETUP

The SETUP specifies the transport mechanism to be used for the streamed media. A receiver shall indicate the transport parameters even if it has no influence over these parameters, for example, in cases where the server advertises a fixed multicast address.

SETUP <URI> RTSP/<Version>

```
ex.  C->S:  SETUP rtsp://live.example.com/concert/audio RTSP/1.0
      CSeq: 1
      Transport: RTP/AVP;unicast;client_port=4588-4589
      S->C:  RTSP/1.0 200 OK
      CSeq 1
      Session: 47112344
      Transport: RTP/AVP;unicast;server_port=6256-6257
```

6.3.2 PLAY

The PLAY method tells the server to start sending data via the mechanism specified in SETUP. A receiver shall not send a PLAY request until any outstanding SETUP requests have been acknowledged as successful.

The PLAY request positions the normal play time to the beginning of the range specified and delivers stream data until the end of the range is reached. PLAY requests may be queued. A server

shall queue PLAY requests to be executed in order. That is, a PLAY request arriving while a previous PLAY request is still active, is delayed until the first request has been completed.

PLAY <URI> RTSP/<Version>

```
ex.  C->S:  PLAY rtsp://live.example.com/concert/audio RTSP/1.0
      CSeq: 2
      Session: 47112344
      S->C:  RTSP/1.0 200 OK
      CSeq 2
```

6.3.3 PAUSE

The PAUSE method causes the stream delivery to be interrupted temporarily. If the request URL names a group of streams, delivery of all currently active streams within the group is halted. After resuming playback, synchronization of the tracks shall be maintained.

PAUSE <URI> RTSP/<Version>

```
ex.  C->S:  PAUSE rtsp://live.example.com/concert/audio RTSP/1.0
      CSeq: 3
      Session: 47112344
      S->C:  RTSP/1.0 200 OK
      CSeq: 3
```

6.3.4 TEARDOWN

The TEARDOWN request stops the stream delivery for the given URI, freeing the resources associated with it. After a TEARDOWN, any RTSP session associated with the URI is no longer valid. A SETUP request has to be issued before the session can be played again.

TEARDOWN <URI> RTSP/<Version>

```
ex.  C->S:  TEARDOWN rtsp://live.example.com/concert RTSP/1.0
      CSeq: 4
      Session: 47112344
      S->C:  RTSP/1.0 200 OK
      CSeq: 4
```

6.4 RTSP header fields

6.4.1 CSeq

The CSeq field specifies the sequence number for RTSP request-response pair. The field shall be present in all requests and responses. For every RTSP request containing the given sequence number, there will be a corresponding response having the same number. Any retransmitted request shall contain the same sequence number as the original (i.e. the sequence number is not incremented for retransmissions of the same request).

6.4.2 Range

The request and response header field specifies a range of time. The range can be specified in a number of units. The specification defines the SMPTE, NPT and clock range units. The header may also contain a time parameter in UTC, specifying the time at which the operation is to be made effective. Range units are summarized in Table 4.

Table 4/J.120 – Formats of the range units

Range units	Formats
SMPTE	hours:minutes:seconds:frames.subframes
NPT	Seconds hours:minutes:seconds now
Clock	<YYYYMMDD>T<HHMMSS.fraction>

6.4.3 Session

This request and response header field identifies an RTSP session started by the media server in a SETUP response and concluded by TEARDOWN. The session identifier is chosen by the media server. Once a receiver obtains a Session identifier, the receiver shall return it for any request related to that session.

6.4.4 Transport

This request header indicates which transport protocol is to be used and configures its parameters such as destination address, compression, multicast time-to-live and destination port for a single stream. It sets those values not already determined by a presentation description. The arguments are separated by ";" and are listed as below. See Table 5.

Table 5/J.120 – The arguments of the Transport

Arguments	Description
<transport-spec>	<transport-protocol>/<profile>[/lower-transport] This value is fixed as "RTP/AVP".
<delivery type>	"unicast" or "multicast". Default value is "multicast".
Destination	The address to which a stream will be sent for multicast.
Ttl	Multicast time-to-live.
Port	This parameter provides the RTP/RTCP port pair for multicast session. It is specified as a range, e.g. port=3456-3457
client_port	This parameter provides the unicast RTP/RTCP port pair on the receiver. It is specified as a range.
server_port	This parameter provides the unicast RTP/RTCP port pair on the server. It is specified as a range.

7 RTP header formats

The RTP header has the following format, see Figure 2.

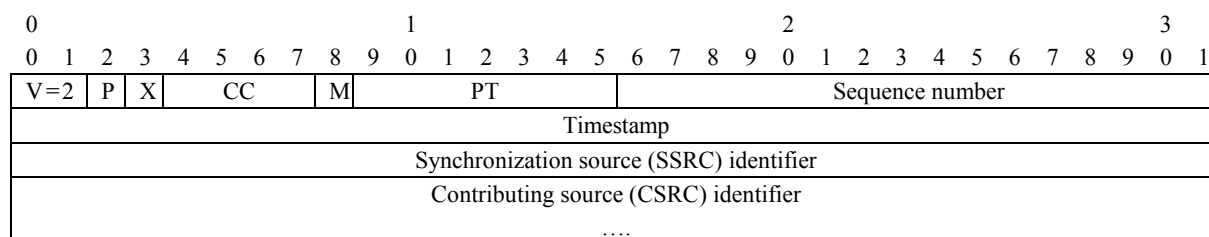


Figure 2/J.120 – RTP header fields

The first twelve octets are present in every RTP packet, while the list of CSRC identifiers is present only when inserted by a mixer. The fields have the following meaning:

Version (V): 2 bits

This field identifies the version of RTP. The version defined by ITU-T H.225.0 is 2. The value 1 is used by the first draft version of RTP and the value 0 is used by the protocol initially implemented in the "vat" audio tool.

Padding (P): 1 bit

If the padding bit is set, the packet contains one or more additional padding octets at the end which are not part of the payload. The last octet of the padding contains a count of how many padding octets should be ignored. Padding may be needed by some encryption algorithms with fixed block sizes or for carrying several RTP packets in a lower-layer protocol data unit.

Extension (X): 1 bit

If the extension bit is set, the fixed header is followed by exactly one header extension, with a format defined in A.5.3/H.225.0.

CSRC count (CC): 4 bits

The CSRC count contains the number of CSRC identifiers that follow the fixed header.

Marker (M): 1 bit

The interpretation of the marker is defined by a profile. It is intended to allow significant events such as frame boundaries to be marked in the packet stream. A profile may define additional marker bits or specify that there is no marker bit by changing the number of bits in the payload type field.

Payload type (PT): 7 bits

This field identifies the format of the RTP payload and determines its interpretation by the application. A profile specifies a default static mapping of payload type codes to payload formats. Additional payload type codes may be defined dynamically through a non-RTP means. An initial set of default mappings for audio and video is specified in ITU-T H.225.0.

Sequence Number: 16 bits

The sequence number increments by one for each RTP data packet sent and may be used by the receiver to detect packet loss and to restore the packet sequence. The initial value of the sequence number is random (unpredictable) to make known-plain text attacks on encryption more difficult, even if the source itself does not encrypt, because the packets may flow through a translator that does encrypt.

Timestamp: 32 bits

The timestamp reflects the sampling instant of the first octet in the RTP data packet. The sampling instant shall be derived from a clock that increments monotonically and linearly in time to allow synchronization and jitter calculations. The resolution of the clock shall be sufficient for the desired synchronization accuracy and for measuring packet arrival jitter (one tick per video frame is typically not sufficient). The clock frequency is dependent on the format of data carried as a payload and is specified statically in the profile or payload format specification that defines the format, or may be specified dynamically for payload formats defined through non-RTP means. If RTP packets are generated periodically, the nominal sampling instant as determined from the sampling clock is to be used, and not a reading of the system clock. As an example, for fixed-rate audio the timestamp clock would likely increment by one for each sampling period. If an audio application reads blocks covering 160 sampling periods from the input device, the timestamp would be increased by 160 for each such block, regardless of whether the block is transmitted in a packet or dropped as silent.

The initial value of the timestamp is random, just the same as for the sequence number. Several consecutive RTP packets may have equal timestamps if they are (logically) generated at once, e.g. belong to the same video frame. Consecutive RTP packets may contain timestamps that are not monotonic if the data is not transmitted in the order it was sampled, as in the case of MPEG interpolated video frames. The sequence numbers of the packets as transmitted will still be monotonic.

SSRC: 32 bits

The SSRC field identifies the synchronization source. This identifier is chosen randomly, with the intent that no two synchronization sources within the same RTP session will have the same SSRC identifier. Although the probability of multiple sources choosing the same identifier is low, all RTP implementations shall be prepared to detect and resolve collisions. If a source changes its source transport address, it shall also choose a new SSRC identifier to avoid being interpreted as a looped source.

CSRC list (Optional): 0 to 15 items, 32 bits each

The CSRC list identifies the contributing sources for the payload contained in this packet. The number of identifiers is given by the CC field. If there are more than 15 contributing sources, only 15 may be identified. CSRC identifiers are inserted by mixers, using the SSRC identifiers of contributing sources. For example, for audio packets the SSRC identifiers of all sources that were mixed together to create a packet are listed, allowing correct talker indication at the receiver.

APPENDIX I

Simple example of RTSP presentation description and session control

This is an example of live program distribution.

C: Client

W: Web server

M: Media server

`http://www.example.com/live.sdp` (Session Description Protocol file)

C->W: GET /live.sdp HTTP/1.0

W->C: HTTP/1.0 200 OK

Content-Type: application/sdp

Content-Length: 44

v=0

o=ITU 2890844526 2890842807 IN IP4 192.16.24.202

s=RTSP Session

i=Live concert

m=audio 3456 RTP/AVP 4 (4=G.723)

a=control:rtsp://live.example.com/concert/audio

/* Download site */

a=decoder:ftp://ftp.example.com/decoder/audio.exe

m=video 3458 RTP/AVP 34 (34=H.263)

a=control:rtsp://live.example.com/concert/video

/* Download site */

a=decoder:ftp://ftp.example.com/decoder/video.exe

C->M: SETUP rtsp://live.example.com/concert/audio RTSP/1.0

CSeq: 1

Transport: RTP/AVP;multicast

```

M->C: RTSP/1.0 200 OK
      CSeq: 1
      Session: 50887676
      Transport: RTP/AVP;multicast;destination=224.0.1.11;
                port=3456-3457;ttl=127

C->M: SETUP rtsp://live.example.com/concert/video RTSP/1.0
      CSeq: 2
      Session: 50887676
      Transport: RTP/AVP;multicast

M->C: RTSP/1.0 200 OK
      CSeq: 2
      Transport: RTP/AVP;multicast;destination=224.0.1.12;
                port=3458-3459;ttl=127

C->M: PLAY rtsp://live.example.com/concert RTSP/1.0
      CSeq: 3
      Session: 50887676

M->C: RTSP/1.0 200 OK
      CSeq: 3

C->M: TEARDOWN rtsp://live.example.com/concert RTSP/1.0
      CSeq: 4
      Session: 50887676

M->C: RTSP/1.0 200 OK
      CSeq: 4

```

APPENDIX II

A typical RTP payload format for ITU-T H.263

An RTP payload format for Webcasting per ITU-T H.263 is described in Figure II.1.

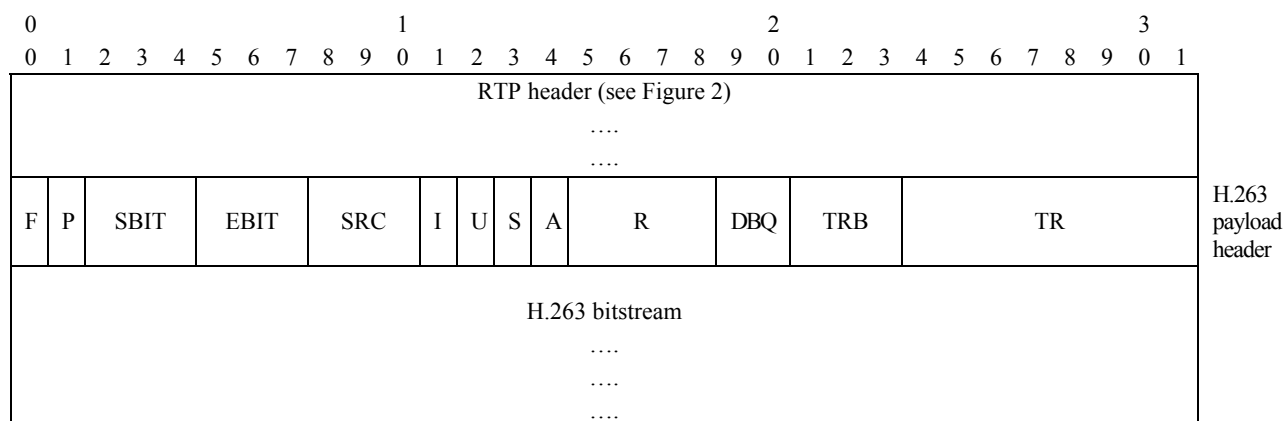


Figure II.1/J.120 – RTP payload format for H.263

In RFC 2190, an RTP Payload format for very low bit-rate video streams, ITU-T H.263, is specified. Three modes are defined for the H.263 payload header. An RTP packet can use one of the three modes for H.263 video streams depending on the desired network packet size and the H.263 encoding options employed. In this example, only mode "A" which has the shortest header is described, and the others are explained in RFC 2190.

For each RTP packet, the RTP fixed header is followed by the H.263 payload header, which is followed by the standard H.263 video bitstream. The layout of an RTP H.263 video packet (mode A) is shown in Figure II.1. For H.263 video streams, each RTP packet carries only one H.263 video packet. The H.263 payload header is always present for each H.263 video packet.

In mode A, an H.263 bitstream will be packetized on a GOB boundary or a picture boundary. Mode A packets always start with the H.263 picture start code or a GOB, but do not necessarily contain complete GOBs. Four bytes are used for the mode A H.263 payload header. The H.263 payload header definition for mode A is shown in Figure II.1 with flag F = 0. Packets are allowed to start at a GOB boundary even if no GOB header is present in the bitstream for the GOB.

The fields of the payload header have the following meaning:

F: 1 bit

The flag bit indicates the mode of the payload header. F = 0, mode A; F = 1, mode B or mode C depending on the P bit defined below.

P: 1 bit

Optional PB-frames mode as defined by ITU-T H.263. "0" implies normal I or P frame, "1" PB-frames. When F = 1, P also indicates the following modes: mode B if P = 0, mode C if P = 1.

SBIT: 3 bits

Start bit position specifies the number of most significant bits that shall be ignored in the first data byte.

EBIT: 3 bits

End bit position specifies the number of least significant bits that shall be ignored in the last data byte.

SRC: 3 bits

Source format, bits 6, 7 and 8 in PTYPE defined by ITU-T H.263, specify the resolution of the current picture.

I: 1 bit

Picture coding type, bit 9 in PTYPE defined by ITU-T H.263, "0" is intra-coded, "1" is inter-coded.

U: 1 bit

Set to 1 if the Unrestricted Motion Vector option, bit 10 in PTYPE defined by ITU-T H.263 was set to 1 in the current picture header, otherwise set to 0.

S: 1 bit

Set to 1 if the Syntax-based Arithmetic Coding option, bit 11 in PTYPE defined by ITU-T H.263 was set to 1 for the current picture header, otherwise set to 0.

A: 1 bit

Set to 1 if the Advanced Prediction option, bit 12 in PTYPE defined by ITU-T H.263 was set to 1 for the current picture header, otherwise set to 0.

R: 4 bits

Reserved, shall be set to zero.

DBQ: 2 bits

Differential quantization parameter used to calculate quantizer for the B frame based on quantizer for the P frame, when PB-frames option is used. The value should be the same as DBQUANT defined by ITU-T H.263. Set to zero if the PB-frames option is not used.

TRB: 3 bits

Temporal Reference for the B frame as defined by ITU-T H.263. Set to zero if the PB-frames option is not used.

TR: 8 bits

Temporal Reference for the P frame as defined by ITU-T H.263.

APPENDIX III

An example of RTP payload format for audio data

RTP Payload format for audio data is specified in RFC 2198. This format is designed under encoding-independent rules. An RTP packet containing audio data should have a standard RTP header with payload type (PT) indicating the type of encoding. Other fields of the RTP header relate to the primary data block of the audio data.

Following the RTP header are a number of additional headers defined in the figure below, which specify the content of each of the encodings carried by the packet. Following these additional headers are a number of data blocks, which contain the standard RTP payload data for these encodings. Note that all the headers are aligned to a 32 bit boundary, but that the payload data will typically not be aligned.

An example of an RTP audio data packet is illustrated, see Figure III.1.

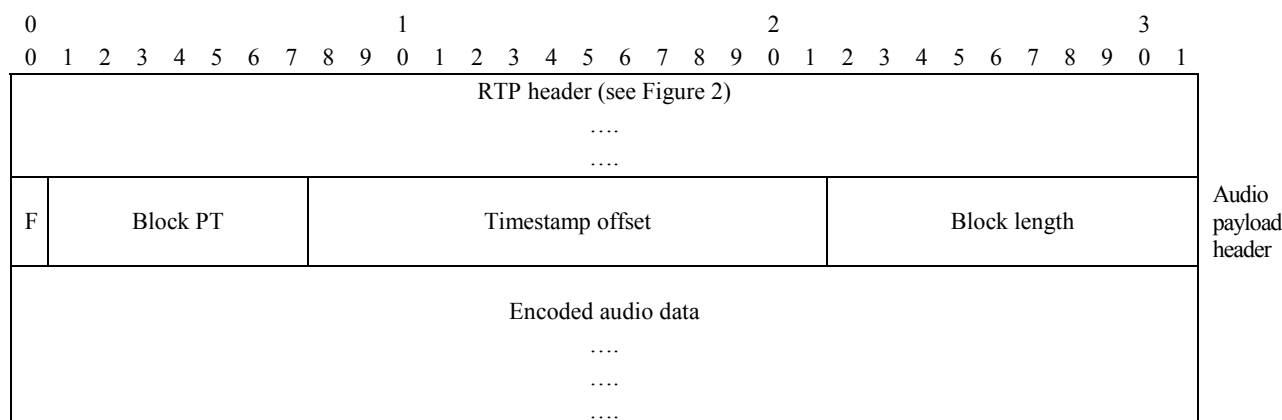


Figure III.1/J.120 – RTP payload format for audio data

The bits in the header are specified as follows:

F: 1 bit

First bit in header indicates whether another header block follows. If this is 1, further header blocks follow, if this is 0, then this is the last header block.

Block PT: 7 bits

7 bits RTP payload type for this block.

Timestamp offset: 14 bits

An unsigned offset of the timestamp of this block relative to the timestamp given in the RTP header. The use of an unsigned offset implies that redundant data shall be sent after the primary data, and is hence a time to be subtracted from the current timestamp to determine the timestamp of the data for which this block is the redundancy.

Block length: 10 bits

Length in bytes of the corresponding data block excluding the header.

APPENDIX IV

Preferred embodiments of downloading decoders

As to downloading decoders, it may be a direct transfer of decoder programs via HTTP or FTP, or may be linked to the HTML page that instructs the end-users which program should be downloaded.

If the decoder program is designed to be dependent on a CPU, for example a CPU type, clock frequency and so on, it is necessary that the end-user be able to choose the most optimum decoder program. A typical downloading page is shown in Figure IV.1.

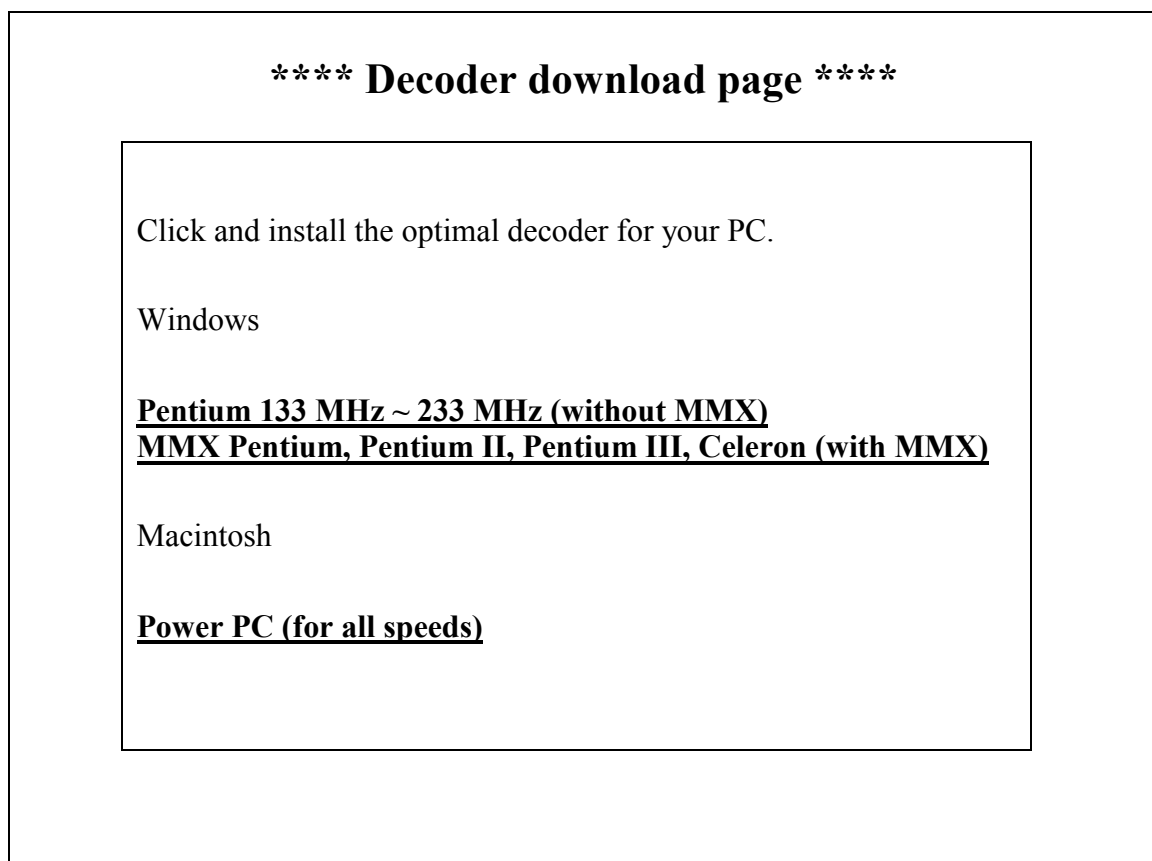


Figure IV.1/J.120 – Typical decoder downloading page

APPENDIX V

Sample Software

Sample source code and some tools are available as an electronic attachment to this Recommendation. Table V.1 lists the contents included in this package.

The sample source code is given under Copyright of KDD R&D Laboratories Inc., for the purpose of establishing conformance with J.120 only. The software is not meant for commercial use. Neither the authors nor KDD R&D Laboratories Inc. are held responsible for any defects of this software. There is absolutely no warranty for this software.

Names terminated by "/" are folder names, otherwise filenames.

Table V.1/J.120 – List of files or folders

Name	Description
Client/ Client/Release/ Client.exe PType.tbl Test.sdp	Sample source code (C++) of J.120 client Executable of J.120 client Payload type table for the client program Sample session description file
Server/ Server/Release/ Server.exe Media/ Test_v.rtp Test_a.rtp	Sample source code (C++) of J.120 server Executable of J.120 server Sample media file folder Sample video RTP packet file for the server Sample audio RTP packet file for the server
Lib/ PT04Dec.dll PT34Dec.dll	Decoder dynamic link libraries Sample decoder DLL for payload type 04 (G.723.1) Sample decoder DLL for payload type 34 (H.263)
Tools/ RTPEncoder.exe RTPDecoder.exe	Miscellaneous tools Video (H.263) and audio (G.723.1) encoder which generates RTP files Simple decoder program for checking RTP files generated by the RTPEncoder.exe

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
Series B	Means of expression: definitions, symbols, classification
Series C	General telecommunication statistics
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	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	TMN and network maintenance: international transmission systems, telephone circuits, telegraphy, facsimile and leased circuits
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	Telephone transmission quality, telephone installations, local line networks
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 and open system communications
Series Y	Global information infrastructure and Internet protocol aspects
Series Z	Languages and general software aspects for telecommunication systems

20002