INTERNATIONAL TELECOMMUNICATION UNION

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

J.123
(07/2002)

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

Interactive systems for digital television distribution

Multiplexing format for webcasting on the TCP/IP network

ITU-T Recommendation J.123

ITU-T J-SERIES RECOMMENDATIONS

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

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

ITU-T Recommendation J.123

Multiplexing format for webcasting on the TCP/IP network

Summary

This Recommendation provides a multiplexing format appropriate for audio and video transmission by download-based protocol over TCP/IP without any session control protocols between server and client. However, when session control protocols are not used, some necessary information should be added to the media data. Therefore, this Recommendation defines a multiplexing format particular to webcasting on the TCP/IP. This format carries metadata, digital rights management (DRM) information and formatted text as well as audio and video bitstreams.

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 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.

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.

CONTENTS

ITU-T Recommendation J.123

Multiplexing format for webcasting on the TCP/IP network

1   Scope

This Recommendation defines a multiplexing format appropriate for audio and video transmission by download-based protocol over TCP/IP. By this format, webcasting on TCP/IP network is realized without any session control protocols between server and client.

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.

2.1   Normative reference

[1]   ISO/IEC 14496-1:2001, *Information technology – Coding of audio-visual objects – Part 1: Systems*.

2.2   Informative references

[2]   ITU-T Recommendation J.120 (2000), *Distribution of sound and television programs over the IP network*.

[3]   ISO/IEC 14496-2:2001, *Information technology – Coding of audio-visual objects – Part 2: Visual*.

[4]   ISO/IEC 13818-3:1998, *Information technology – Generic coding of moving pictures and associated audio information – Part 3: Audio*.

[5]   IETF RFC 2068 (1997), *Hypertext Transfer Protocol – HTTP/1.1*.

3   Terms and definitions

This Recommendation defines the following terms:

3.1   box: An object-oriented building block defined by a unique type identifier and length (called 'atom' in the reference [1]).

3.2   chunk: A contiguous set of samples for one track.

3.3   container box: A box whose sole purpose is to contain and group a set of related boxes.

3.4   movie box: A container box whose sub-boxes define the metadata for a presentation ('moov').

3.5   media data box: A container box which can hold the actual media data for a presentation ('mdat').

3.6   presentation: One or more motion sequences, possibly combined with audio.

3.7   sample: An individual frame of video, or a time-contiguous compressed section of audio.

3.8 **sample description**: A structure which defines and describes the format of some number of samples in a track.

3.9 **sample table**: A packed directory for the timing and physical layout of the samples in a track.

3.10 **track**: A collection of related samples, which corresponds to a sequence of images or sampled audio.

3.11 **webcasting**: Webcasting is defined in ITU-T Rec. J.120: Distribution of sound and television programs over the IP network.

4 Abbreviations

This Recommendation uses the following abbreviations:

DRM   Digital Rights Management

HTTP   Hypertext Transport Protocol

IP   Internet Protocol

TCP   Transmission Control Protocol

UUID   Universal Unique Identifier

5 Reference architecture

This Recommendation assumes that download-based protocol (e.g. HTTP) should be used for webcasting because it does not require any complex server-client protocols.

The reference architecture for webcasting on TCP/IP is shown in Figure 1.
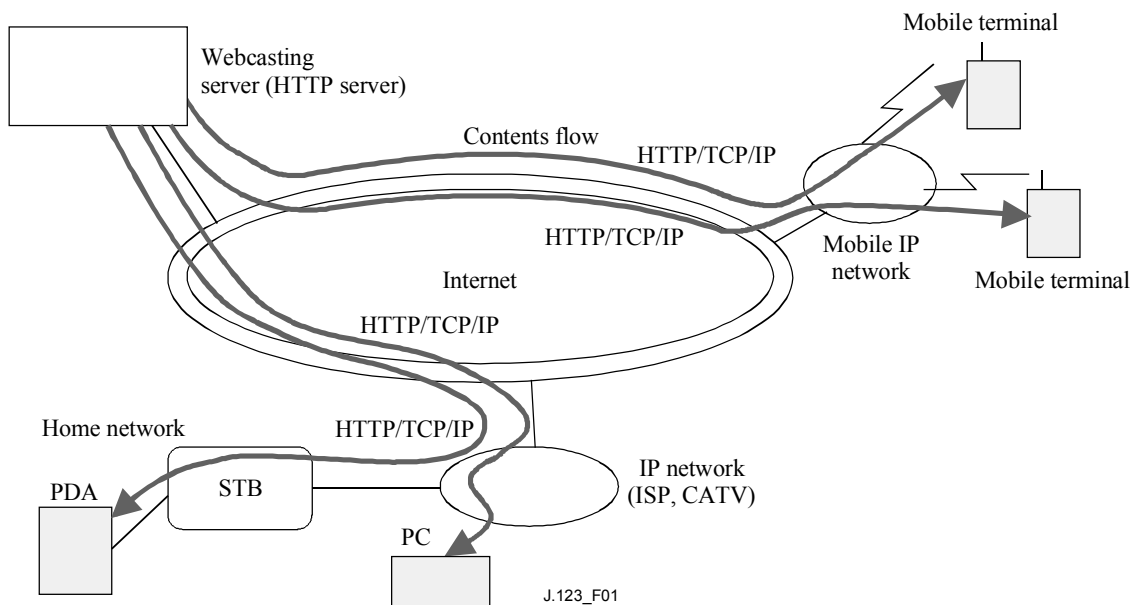
Figure 1/J.123 – Architecture of webcasting on TCP/IP

6 File format

6.1 Basic structure

The file format consists of extension data, contents header and media data. Basic structure of the file is shown in Figure 2.
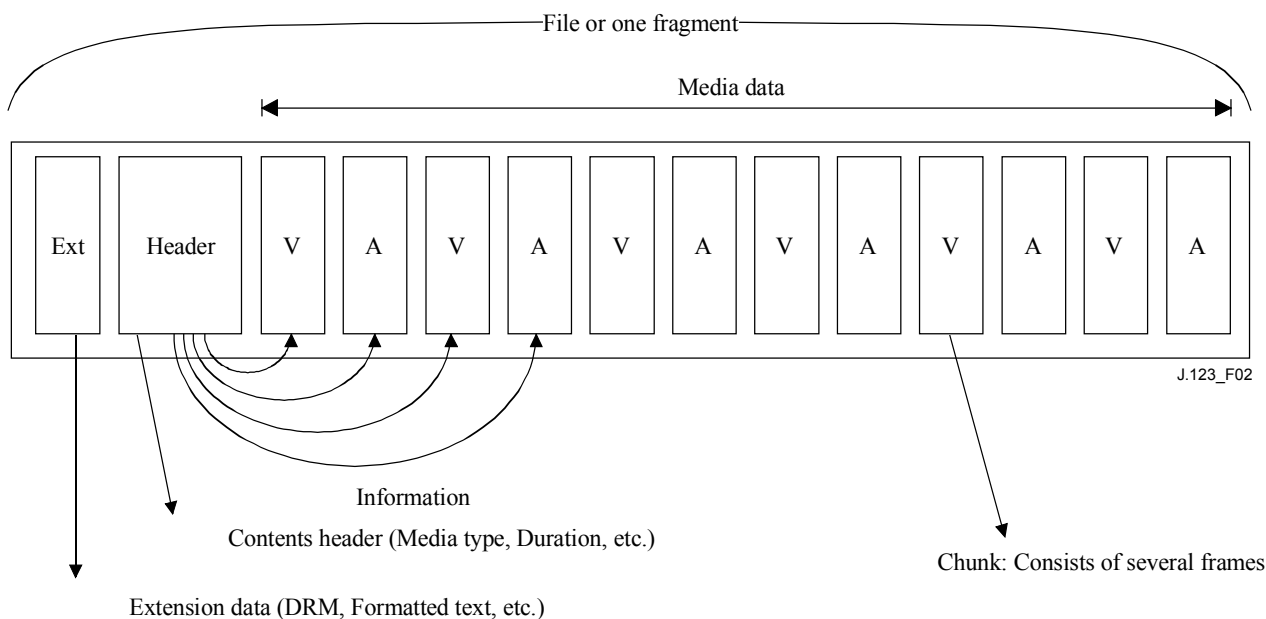
Figure 2/J.123 – Basic structure of a file format

6.2 Object Structure

The file is structured as a sequence of objects called "Box"; some of these objects may contain other objects. The sequence of objects in the file shall contain exactly one presentation metadata wrapper (the Movie Box). It shall be located in the beginning of the file. The other objects found at this level may be UUID Boxes, or Media Data Boxes.

The fields in the objects are stored with the most significant byte first, commonly known as network byte order or big-endian format.

All boxes defined in this Recommendation are listed in Table 1, which are marked by a grey shade.

Table 1/J.123 – Box types and structure

ftyp				File type and compatibility
uuid				Extension data
moov				Container for all the information
	mvhd			Movie header, overall declarations
	trak			Container for an individual track or stream
		tkhd		Track header, overall information about the track
		tref		Track reference container
		edts		Edit list container
			elst	An edit list
		mdia		Container for the media information in a track
			mdhd	Media header, overall information about the media
			hdlr	Handler, declares the media (handler) type
			minf	Media information container
				vmhd	Video media header, overall information
				smhd	Sound media header, overall information
				hmhd	Hint media header, overall information
				dinf	Data information box, container
					dref	Data reference box, declares source(s) of media data in track
				stbl	Sample table box, container for the time/space map
					stsd	Sample descriptions (codec types, initialization etc.)
					stts	(Decoding) time-to-sample
					ctts	(Composition) time to sample
					stsc	Sample-to-chunk, partial data-offset information
					stsz	Sample sizes (framing)
					stz2	Compact sample sizes (framing)
					stco	Chunk offset, partial data-offset information
					stss	Sync sample table (random access points)
					stsh	Shadow sync sample table
					padb	Sample padding bits
					stdp	Sample degradation priority
	mvex			Movie extends box
		trex		Track extends defaults
moof				Movie fragment
	mfhd			Movie fragment header
	traf			Track fragment
		tfhd		Track fragment header
		trun		Track fragment run
mdat				Media data container
free				Free space
skip				Free space
	udta			User-data

7 Box definitions

7.1 File Type box

7.1.1 Definition

Box Type: 'ftyp'

Container: File

Mandatory: Yes

Quantity: Exactly one

A media-file structured to this part of this specification may be compatible with more than one detailed specification, and it is therefore not always possible to speak of a single 'type' or 'brand' for the file. This means that the utility of the file name extension and mime type are somewhat reduced.

This box must be placed as early as possible in the file (e.g. after any obligatory signature, but before any significant variable-size boxes such as the UUID Box, Movie Box or Media Data Box). It identifies which specification is the 'best use' of the file, and a minor version of that specification; and also a set of other specifications to which the file complies. Readers implementing this format should attempt to read files which are marked as compatible with any of the specifications which the reader implements. Any incompatible change in a specification should therefore register a new 'brand' identifier to identify files conformant to the new specification.

The type 'isom' is defined in this clause, as identifying files which conform to the format in this Recommendation. More specific identifiers can be used to identify precise versions of specifications providing more detail.

Files would normally be externally identified (e.g. with a file extension or mime type) that identifies the 'best use' (major brand), or the brand that the author believes will provide the greatest compatibility.

7.1.2 Syntax

```
aligned(8) class FileTypeBox
    extends Box('ftyp') {
    unsigned int(32)   major-brand;
    unsigned int(32)   minor-version;
    unsigned int(32)   compatible-brands[];   // to end of the box
}
```

7.1.3 Semantics

This box identifies the specifications to which this file complies.

Each brand is a printable four-character code that identifies a precise specification. Only one brand is defined here: 'isom', identifies files structurally conformant to this media-independent part of this specification.

major-brand – is a brand identifier

minor-version – is an informative integer for the minor version of the major brand

compatible-brands – is a list, to the end of the box, of brands

7.2 Other boxes

Definitions of all other boxes are found in the reference [1].

8 Extension Data

Extension Data is formatted in "uuid" box. Two functions for Extension Data are described as follows:

- Digital Rights Management
    - Copy prohibition
    - Expiration date
    - Validation period after downloading
    - Number of times play
- Formatted Text (Closed Caption)
    - Text decoration: Foreground/background colour, Bold, Italic, Fonts, etc.
    - Effects: Scroll, Wipe, Blink, etc.
    - Synchronization with the media data
    - Hyper link to the Internet

8.1 Digital Rights Management

Rights management information controls play and/or re-transmission of the downloaded MP4 file. It is contained in uuid Box of the file format.

8.1.1 Syntax

```
aligned(8) class CopyGuardBox extends FullAtom ('uuid', version = 0, flags){
    bit(32)            copy-guard;
    unsigned int(32)   limit-date;
    unsigned int(32)   limit-period;
    unsigned int(32)   limit-count;
}
```

8.1.2 Semantics

Field	Type	Description	Parameters
type	uint32	Type of Box	'uuid' is set
usertype	uint8[16]	ID	"cpgd"-A88C-11d4-8197-09027087703
version	uint8	Version	0 is set
flags	bit24	Rights management flags	0: No limitation 1: Limitation by expiration date 2: Limitation by validated period 4: Limitation by playing number of times Unless the case of No limitation, the following "never copy" flag shall be set to '1'
copy-guard	bit32	"never copy" flag	0: copy permitted 1: copy prohibited
limit-date	uint32	Expiration date	Specify the expiration date in seconds from 1904/1/1 0:00GMT
limit-period	uint32	Validated period	Specify the validated period in days after the file is downloaded
limit-count	uint32	Playing number of times	'1' means that the file can be played only once

8.2 Formatted text

The formatted text is a closed caption synchronizing to a video content. It is contained in uuid Box of the file format.

8.2.1 Basic structure

The formatted text is written by XML-based syntax. The following is an overview of the elements:

tsml				Root
	head			Header
		layout		Layout
			region	Region attribute for displaying telop
			font	Default font
	body			Telop body
		telop		Telop letters
			font	Font for letters
			br	Line break
			u	Underline
			rev	Reversal of font and background color
			a	Link

Syntax of the formatted text is shown as follows.

```
<tsml>
    <head>
        :
    </head>
    <body>
      <telop> ... </telop>
      <telop> ... </telop>
          :
    </body>
</tsml>
```

In this Recommendation, XML declaration is omitted.

8.2.2 Text describing and controlling element

Each element is described using the following parameters:

•   attribute

     It is an inherent attribute that is described in the beginning tag.

•   child

     It shows elements that can be included between the beginning tag and the end tag.

8.2.3 <tsml> element

<tsml> element is a root of the formatted text.

Tag		
<tsml> </tsml>		
attribute		
Name	Value (default)	Description
child		
head	header of telop information	
body	body part of telop information	

NOTE – <tsml> element appears only once in the formatted text information.

8.2.4 <head> element

<head> element specifies common settings throughout the formatted text information.

Tag		
<head> </head>		
attribute		
Name	Value (default)	Description
child		
layout	Layout for displaying telop	

8.2.5 <layout> element

<layout> element specifies the attribute of the displaying area.

Tag		
<layout> </layout>		
attribute		
Name	Value (default)	Description
child		
region	attribute of telop region	
font	default font	

8.2.6 <region> element

<region> element specifies the attribute of the formatted text region.

Tag		
<region> </region>		
attribute		
Name	Value (default)	Description
background-color	color value #rrggbb	Background color of the region
child		

8.2.7 element in <head> element

 element specifies the attribute of the font such as a default font color.

Tag		
 		
attribute		
Name	Value (default)	Description
color	color value #rrggbb	font color
child		

8.2.8 <body> element

<body> element includes a body part of the formatted text.

Tag		
<body> </body>		
attribute		
Name	Value (default)	Description
child		
telop	telop letters	

8.2.9 <telop> element

<telop> element specifies appearance and letters of the formatted text.

Tag		
<telop> </telop>	Texts are included between these tags	
attribute		
Name	Value (default)	Description
begin	"X" (0)	begin time of the element in ms
end	"X" (The end time of the content in the MP4 file)	end time of the element in ms
wrap	true | false (false)	automatic word wrap
child		
font	Fonts for letters	
br	Line break	
u	Underline	
rev	Reversal of colours for fonts and background	
a	Links	

NOTE 1 – The beginning time of the contents is taken as 0 ms.

NOTE 2 – If the next <telop> element has earlier beginning time than the end time of the previous <telop>, the next <telop> starts at the beginning time as specified. (The beginning time is given priority.)

The formatted text is displayed at the time of 'begin', and disappeared at the time of 'end'.

Following elements are included in the <telop> element as a child element.

8.2.10 element in <telop> element

Tag		
 		
attribute		
Name	Value (default)	Description
color	COLOR value #rrggbb	Font color
child		
br	Line break	
u	Underline	
rev	Reversal of font color and background color	
a	Link	

8.2.11 <br> element

Tag		
attribute		
Name	Value (default)	Description
child		

This element inserts a line break at the arbitrary letter.

8.2.12 <u> element

Tag		
<u></u>		
attribute		
Name	Value (default)	Description
child		
font	Font color	
br	Line break	
rev	Reversal of font color and background color	
a	Link	

Enclosed letters by this element are given an underline.

8.2.13 <rev> element

Tag		
<rev></rev>		
attribute		
Name	Value (default)	Description
child		
font	Font color	
br	Line break	
u	Underline	
a	Link	

Reversal of font color and background color.

8.2.14 <a> element

Tag		
<a>		
attribute		
Name	Value (default)	Description
href	Link destination	
child		
font	Font color	
br	Line break	
u	Underline	
rev	Reversal of font color and background color	

Link destination can include telephone number (tel:), e-mail address (mailto:), and WWW (http:).

8.2.15 Limitations for nesting of elements

<font>, <u> and <rev> element are permitted only one nest.

```
ex.
<font><u> ... </u></font>          OK
<font><u><font> ... </font></u></font>     NG
<font><rev><u> ... </u></rev></font>  NG
```

Appendix I

Sample parameters

Video encoding type: ISO/IEC MPEG-4 Visual Simple Profile Level 1

Audio encoding type: ISO/IEC MPEG-2 Audio Layer 3 (22 050 Hz)
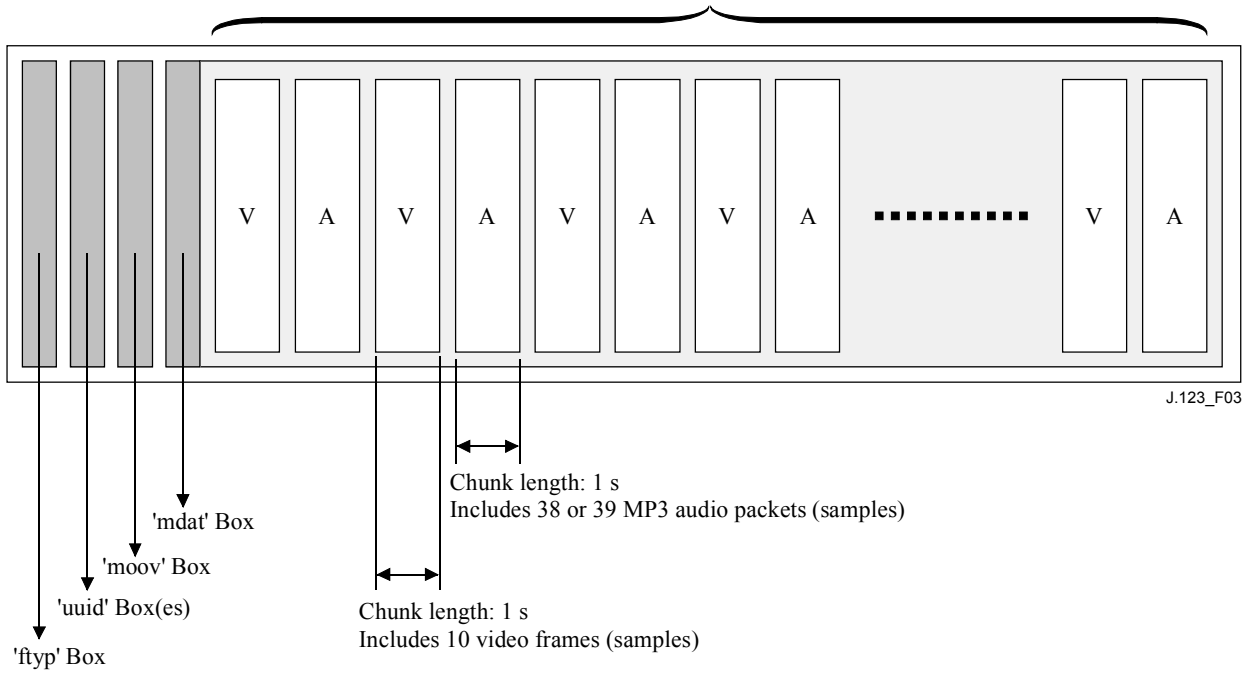
Content length: 30 seconds

Video frame rate: 10 frames/second

Interleave interval (chunk length):   1 second

Multiplexing by these parameters results in the following format:

30 video chunks and 30 audio chunks

Chunk length: 1 s
Includes 38 or 39 MP3 audio packets (samples)

Chunk length: 1 s
Includes 10 video frames (samples)

'mdat' Box

'moov' Box

'uuid' Box(es)

'ftyp' Box

J.123_F03

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   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   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