Technologies and Services on Digital Broadcasting (5)

Overview of MPEG-2 Systems

Part 1 of the MPEG-2 Standard(1) called "Systems" (MPEG-2 Systems) specifies a systems coding for the construction of audio visual systems using video and audio bitstreams compressed and coded by the specifications prescribed in Part 2 "Video" and Part 3 "Audio."
The MPEG-1 Systems Standard, which became an International Standard (IS) in 1993, assumes storage systems like CD-Interactive (CD-I) and specifies a standard that presumes an error-free transmission channel. In contrast, the MPEG-2 Systems Standard takes into account not only storage systems but also telecommunications systems like Asynchronous Transfer Mode (ATM) and broadcasting systems like multi-channel digital broadcasting. It is a comprehensive standard that considers environments in which errors can occur in the transmission channel. The MPEG-2 Systems Standard can be said to be the first international standard for the mutual exchange of multimedia information between various types of information systems, including those for communications, storage, and broadcasting. Despite this comprehensiveness, additional specifications may be needed when developing specific application systems. For example, program-scheduling information and data broadcasting schemes in a digital broadcasting system have yet to be specified. This section describes the role of MPEG-2 Systems in digital broadcasting as well as the basic technical concepts and specific stipulations of this standard.

1. Role of MPEG-2 Systems

The MPEG-2 Systems Standard specifies a system for combining one or more MPEG video and audio data streams so that video and audio, compressed and coded according to the MPEG standard, can be used for a variety of applications. Figure 1 outlines the MPEG-2 Systems Standard. The process shown here converts multiple bit streams (referred to below as "elementary streams"), which are obtained by compressing and coding the video and audio of one or more programs into one multiplexed stream. At the decoder, after transmission of the stream over possibly many different transmission channels (including storage systems), the process extracts the elementary streams making up the target programs and plays them back while preserving the temporal synchronization among individual elementary streams.

In this way, the objective of the MPEG-2 Systems standard is to provide coding for multiplexing, transmitting, and demultiplexing multiple programs and coding to describe information for achieving synchronization between elementary streams.

The following points must be kept in mind when incorporating the MPEG-2 Systems Standard into a broadcasting system.

(1) The MPEG-2 Systems specification is mainly a project for multiplexing MPEG video and audio bit streams.
(2) A standard for transmitting bitstreams of information for the purpose of broadcasting JPEG still pictures and other data is not included and is specified separately in the broadcasting system.
(3) While minimal scrambling-control functions are included, such as for turning scrambling on and off and updating keys, scrambling management information is specified separately in the broadcasting system.
(4) MPEG-2 Systems is a standard specifying the signal input to the decoder. It does not specify, for example, the system for controlling the amount of delay in each video and audio signal at the time of multiplexing.

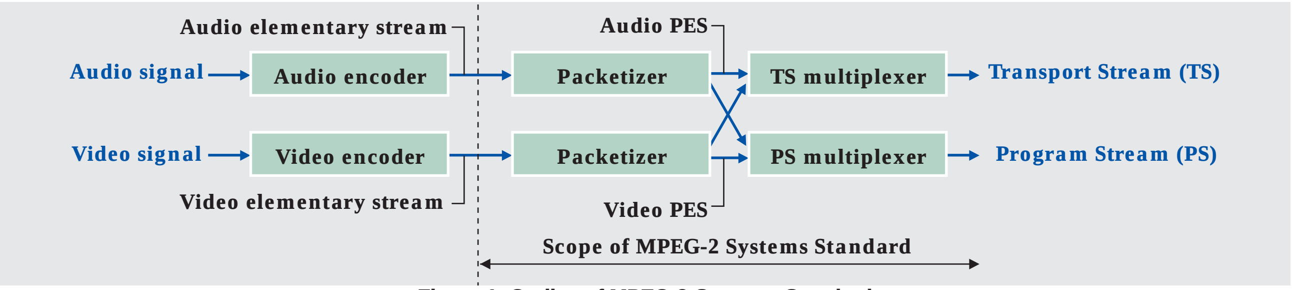


Figure 1: Outline of MPEG-2 Systems Standard

Table 1: Requirements

Item	Description
(1) Uniform syntax	Uniform coding covering maximum functions
(2) Audio and video multiplexing	Multiplexing function for multilayer encoded audio and video
(3) Robust against bit error and cell loss	Recover function for handling bit error and lost packets
(4) Backward compatibility with MPEG1 and H.261	MPEG2 program stream decoder shall eventually be able to decode MPEG1 program streams. In addition, coding shall allow for transparent transmission of MPEG1 and H.261.
(5) Random access and channel hopping	The time from program or channel selection to display shall be short.
(6) Special (trick) mode	Functions such as fast-forward, freeze, and reverse
(7) Multichannel and multilingual audio	Multiplexing of multichannel and multilingual audio
(8) Multiprogramming	Multiplexing of multiple programs
(9) Data-stream identification signal	Identification of multiplexed video/audio elementary streams
(10) Encryption and scrambling	Scrambling at the PES-packet and transport-packet levels
(11) Editing	It shall be possible to connect different elementary streams with minimum discontinuity.
(12) ATM support	Use on an ATM network shall be taken into account.
(13) Variable bit-rate operation	Multiplexing of variable bit-rate elementary streams
(14) Stream remultiplexing	It shall be possible to reconfigure a single multiplexed stream and create another multiplexed stream
(15) Reconfiguring of video block	At the decoder, it shall be possible to regenerate and use a video clock from reference clock information contained in the multiplexed stream.

2. Requirements

The 15 requirements listed in Table 1 are embodied in the MPEG-2 Systems Standard.

3. Transport Stream and Program Stream

The purpose of system coding is to achieve the desired functions efficiently by adding a certain amount of overhead information to data that is to be transmitted. MPEG-2 Systems adopt a packet multiplexing system to combine video, audio, and data streams. In general, a packet in this system consists of a section called a "packet header," which holds control information for identifying packets and performing other functions, and a subsequent data section. The packet multiplexing system enables various kinds of data to be handled equally in units of packets, thereby making for efficient and flexible multiplexing.

The MPEG-2 Systems Standard specifies two kinds of multiplex stream. One is the Transport Stream (TS) that assumes the use of transmission channels (such as in broadcasting and telecommunications) where errors occur; a stream of this type consists of relatively short fixed-length TS packets. The other is the Program Stream (PS) that assumes the use of error-free transmission channels (storage); a stream of this type consists of relatively long variable-length packets called Packetized Elementary Streams (PES). The above specifications make it possible to convert one type of stream into the other. For example, a PES packet can be divided into segments that can then be placed in the data sections of TS packets. A PS "pack," in contrast, includes multiple PES packets of nearly the same length containing video and audio information

Table 2 summarizes the applications and features of the Transport Stream and Program Stream, and Fig. 2 shows how the Transport Stream, Program Stream, and PES packets are related. The following discussion focuses on the Transport Stream, taking into account its potential application to broadcasting.

Table 2: Applications and features of Transport Stream and Program Stream

	Applications	Features
Transport Stream (TS)	Used for transmission and storage in environments where cell loss and bit errors occur. (Error correction code is not included in the MPEG standard.) Examples: broadcasting, telecommunications (ATM)	Fixed-length packets (188 bytes) A PES packet (which is variable length and long compared with TS-packet length) will be divided up and placed in the payload (data) section of TS packets. Multiple reference clocks possible. (Different reference clocks can be used for different programs.)
Program Stream (PS)	Used for transmission and storage in environments where errors do not occur. Examples: digital storage media (DSM) like DVD and CD-ROM	Consists of multiple packs (variable length) each made up of multiple PES packets. Similar to program stream specified in MPEG-1 Systems. Has a single reference clock.

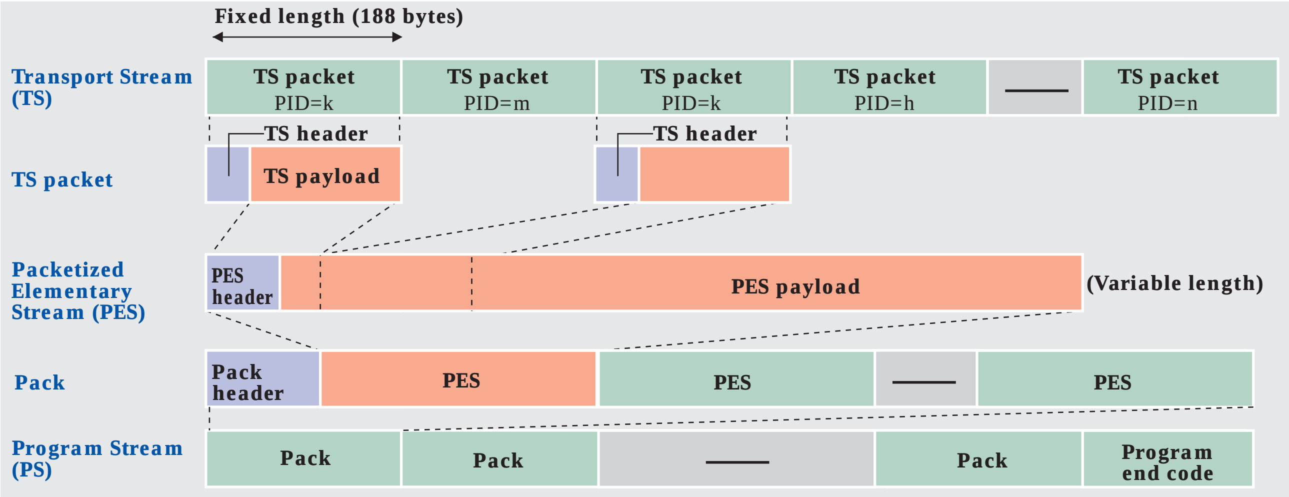


Figure 2: Relationship between Transport Stream, Program Stream, and PES packets

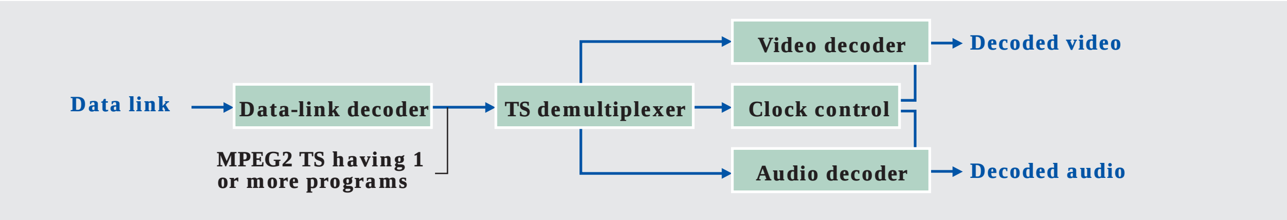
4. Multiplex models

Various types of processing must be performed on the receiver side, such as extracting and decoding a Transport Stream or Program Stream that contains only one program from a multiplex stream that includes multiple programs. Remultiplexing processing at a relay station should also be possible. The MPEG-2 Systems standard specifies the coding scheme that takes the five types of multiplex processing described in Table 3 into account. Figure 3 shows examples of such processing.

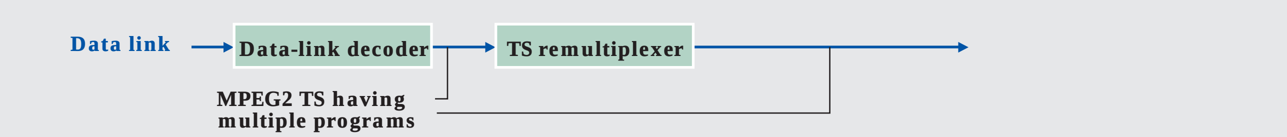
The Transport Stream System Target Decoder (T-STD) consists of the theoretical function blocks that make up the process of decoding multiple elementary streams that make up a single program. In addition, Program Specific

Table 3: Multiplex models

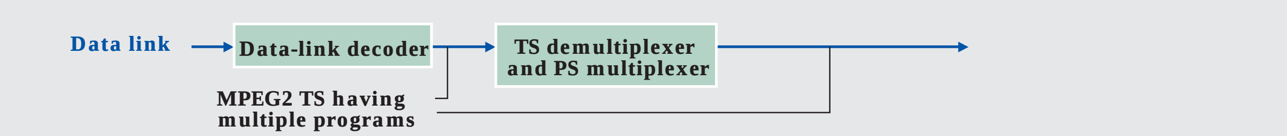
Item	Description	Notes
(1) Demultiplex and decode	Extract a stream consisting of one program from the TS and decode.	See Fig. 3a
(2) Convert to single-program TS	Extract a stream consisting of one program from the TS and output a different TS having that program only.	See Fig. 3b
(3) Convert to multiprogram TS	Extract transport packets consisting of one or more programs from one or more TS's and output a different TS.	Figure omitted
(4)TS→ PS conversion	Extract contents making up 1 program from a TS and output a PS for that program only.	See Fig. 3c
(5) PS→ TS→ PS conversion	Convert a PS to a TS, transmit the TS in an error-prone environment, and output a PS the same as the original PS.	Figure omitted



a: Example of transport demultiplexing and decoding



b: Example of transport multiplexing



c: Conversion from TS to PS

Figure 3: Outline of Transport Stream multiplexing

Information (PSI) supports the selection of multiple elementary streams belonging to one program. The following describes T-STD and PSI.

5. Target-stream system target decoder (T-STD)

To correctly configure a program consisting of multiple streams of video, audio, etc., the constituent elements (video, audio, etc.) of the target program that come to be separated from the multiplex stream must be synchronized for playback. For this reason, the MPEG Systems specification adopts a synchronization control system using time stamps.

(1) Synchronization control system using time stamps [2]

The principle behind this system is as follows. To begin with, the coding side includes a block that feeds a System Time Clock (STC) as a reference time (see Fig. 4). Specifically, the system adds an STC value to video and audio coded data as a time stamp for each unit of presented information and then performs multiplexing of the resulting data. Next, the encoding system inserts a reference clock to regenerate the STC on the receiver side. The receiver side, which also has an STC feed block, regenerates the reference clock inserted in the multiplex stream so that the timing of the decoding process agrees with that of the coding. Here, the system places each unit of coded data in a buffer to generate a delay and decodes and presents the data unit when its time stamp matches the STC. This process corrects the temporal offset between the video and audio streams caused by multiplexing.

(2) System Target Decoder

The System Target Decoder (STD) is a hypothetical decoder model for representing the process of demultiplexing the MPEG multiplex stream, obtaining the individually compressed and coded video and audio streams making up a single program, buffering these streams for the corresponding video and audio decoders, and outputting them in synchronized form. The configuration of an actual system decoder may differ from this model.

The basic operation of the STD is explained here using T-STD, the STD for an MPEG-2 Transport Stream. Figure 5 shows the configuration of T-STD. From the left, T-STD consists of a demultiplexing section, transport buffers, main buffers, and video and audio decoders. In the case of video,

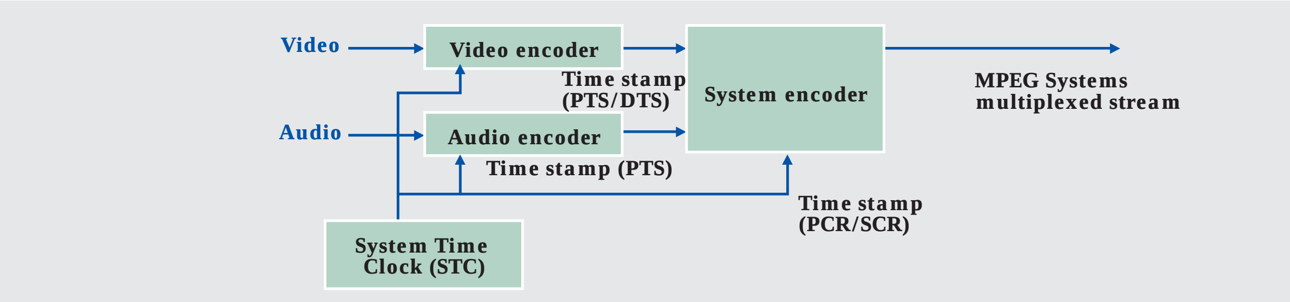
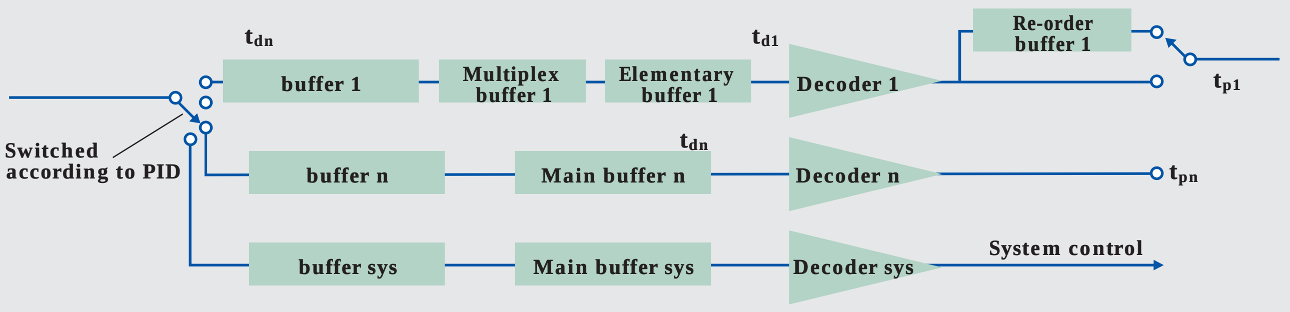


Figure 4: Encoder system example



Re-order buffer: Memory for delaying I/P pictures when B pictures are included.
System control information: Program Specific Information (see section 6).
Decoder: Decodes video and audio data or Program Specific Information.
Main buffer: In the case of audio and data, buffer memory needed to perform delay control to absorb multiplex jitter and to decode audio and data.
Multiplex buffer: In the case of video, buffer memory needed to perform delay control to absorb multiplex jitter.
Elementary-stream buffer: In the case of video, buffer memory for video decoding.
Transport buffer: Preliminary buffer for making high-speed burst-like TS packets compatible with the operating speeds of the main buffer and other subsequent buffers.
t_{dn}: Decoding time for an encoded video frame or for an audio presentation unit having the n-th elementary stream.
t_{pn}: Presentation time of a video frame or audio presentation unit having the n-th elementary stream.

Figure 5: Transport Stream System Target Decoder

the specification calls for the main buffer to be divided into a multiplex buffer and elementary-stream buffer. Also, while not shown in the figure, there is a section that extracts the Program Clock Reference (PCR) for a single program from the bit stream to regenerate the STC.

(3) PTS and DTS

The MPEG-2 System Standard defines two types of time stamp: Presentation Time Stamp (PTS) indicating time of presentation and Decoding Time Stamp (DTS) indicating decoding start time. These time stamps are added to video and audio coding units (corresponding to "picture" for video and "audio frame" for audio) called access units. They are defined as follows for the T-STD of Fig. 5.

To begin with, the specification states that video or audio elementary-stream access units that do not include B pictures defined in MPEG-2 Part 2 Video are to be transferred immediately from the main buffers to the decoders at the time denoted by tpn in Fig. 5 for decoding and display. In other words, this tpn is to be encoded in the field within the PES packet header by the encoder as a PTS. The STD then decodes and outputs the data in the main buffers when the STC matches the PTS.

A video elementary stream that includes B pictures, however, requires that I and P pictures be decoded before decoding the B pictures, and it is for this reason that the decoding time and presentation time of an I or P picture differ. In particular, the specification states that I or P picture data are to be transferred immediately from the main buffer to the decoder at the time denoted by td1 in Fig. 5 for decoding. The encoder encodes this td1 in the field within the PES packet header as a DTS. The system decoder then decodes and outputs I-picture or P-picture data in the main buffer when the STC matches the DTS. After that, a re-order buffer is used to delay display of the I-picture or P-picture signal until the STC matches the PTS.

The PTS and DTS time stamps are expressed in units of one period of a 90-kHz clock and the PCR in units of one period of a 27-MHz clock. The PTS time stamp must be coded at intervals of 0.7 s or less and the PCR within intervals of 0.1 s or less.

(4) Buffer operation

The transport buffer is provided with the aim of buffering the high-speed burst-like data stream that directly follows demultiplexing. It has a capacity of 512 bytes, which is sufficient to store more than two TS packets. Data transfer from the transport buffer to the next buffer is performed on the basis of the leaky bucket model. That is to say, if bytes of data exist in the transport buffer, data bytes will be transferred at a fixed leak rate determined by the type of stream. For MPEG-2 video, for example, the specification calls for 1.2 times the maximum bit rate determined by the profile and level in question. This model simplifies the operation and control of the transport buffers. If no bytes of data exist in the transport buffer, data transfer cannot be performed. The specification also states that the bit stream must be configured so that overflow of the transport buffer does not occur.

The main buffer is divided into a multiplex buffer and an elementary-stream buffer in the case of video, as mentioned above. The function of the multiplex buffer is to buffer 4-ms multiplex jitter and PES overhead and to achieve a data transfer speed that is compatible with the elementary-stream buffer of the next stage. The main buffer is divided into two sections so that this function can be specified in a clear manner. The elementary-stream buffer performs a Video Buffering Verifier (VBV) function as specified in MPEG-2 Part 2 Video.

Two methods, the leak method and the VBV delay method, are specified here for data transfer from the multiplex buffer to the elementary-stream buffer. The leak method employs the leaky bucket model, the same as in the transport buffer. The leak rate here is the maximum bit rate determined by the profile and level of video streams other than the high-1440 or high-level types. The VBV delay method, on the other hand, transfers picture data at a bit rate calculated using the VBV delay value coded within the video stream. (The VBV delay value is the amount of time that one picture must be held in the VBV buffer for decoding. It assumes a fixed end-to-end delay between encoder and decoder buffers.) While the leak method is simpler to implement, it may suffer from a loss of buffer control in trick mode. The VBV delay method should be used in such a case. Which method to employ is determined by whether an STD descriptor exists in the PMT, by examining the leak-valid value and VBV-delay value in the STD descriptor, and by other conditions.

In the case of audio and data, the function of the main buffer is to absorb multiplex jitter and decode the audio and data in question. The capacity of the main buffer is specified as the total amount of memory needed to accommodate a buffer for multiplex jitter, a buffer for overhead in the PES packet header, and a buffer for audio and data decoding.

Based on the above conditions, the job of the main buffer is to store the input stream and to move data on the basis of time stamps. In this regard, the specification states that the bit stream must be configured so that the main buffer does not overflow or underflow while these operations are taking place. Figure 6 shows an example of main buffer operation.

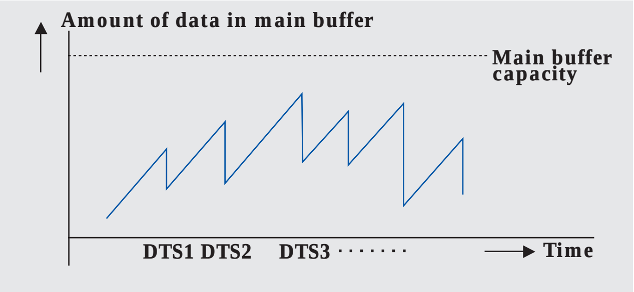


Figure 6: Example of main buffer operation

6. Program Specific Information (PSI)

A Transport Stream includes tables that describe the relationship between the programs that it includes and the elementary streams that make up each of those programs. The information in these tables can be used to indirectly specify the set of packets that configure the program targeted for reception. In particular, a program can be identified as one of about 60,000 types using a 16-bit program number, and the correspondence between this program number and the set of PIDs of TS packets that include the elementary streams of the program can be encoded. This makes it easy to change program configuration and to achieve flexible multiplexing. The information in these tables is called Program Specific Information (PSI). Five tables are specified, as summarized in Table 4. An example of their use is shown in Fig. 7.

In MPEG-2 Systems, "program" specifies the set of elements constituting a program, such as elementary streams. This concept corresponds to "channel" in broadcasting. A broadcast program is called an "event" in MPEG Systems, while MPEG-2 specifies a set of program elements having start and end times within the "program." The "event" identification method is outside the scope of the MPEG-2 Systems standard.

7. Detailed Transport Stream contents

(1) TS packet configuration

A Transport Stream consists of consecutive transport-

Table 4: Program Specific Information (PSI)

Table Name	Specification Type	Allocated PID	Description
Program Association Table	MPEG	0x00	Associates program numbers with PIDs of TS packets sending A Program Map Table
Program Map Table	MPEG	Specified in PAT	Specifies PID values of stream configuring a program
Network Information Table	Private (specified in application	Specified in PAT	Physical network parameters Examples: FDM frequency, transponder number
Conditional Access Table	MPEG	0x01	Associates PID values and EMM stream sending subscription details for charged broadcasting
TS Description Table	MPEG	0x02	Associates descriptors and the entire TS (application system)

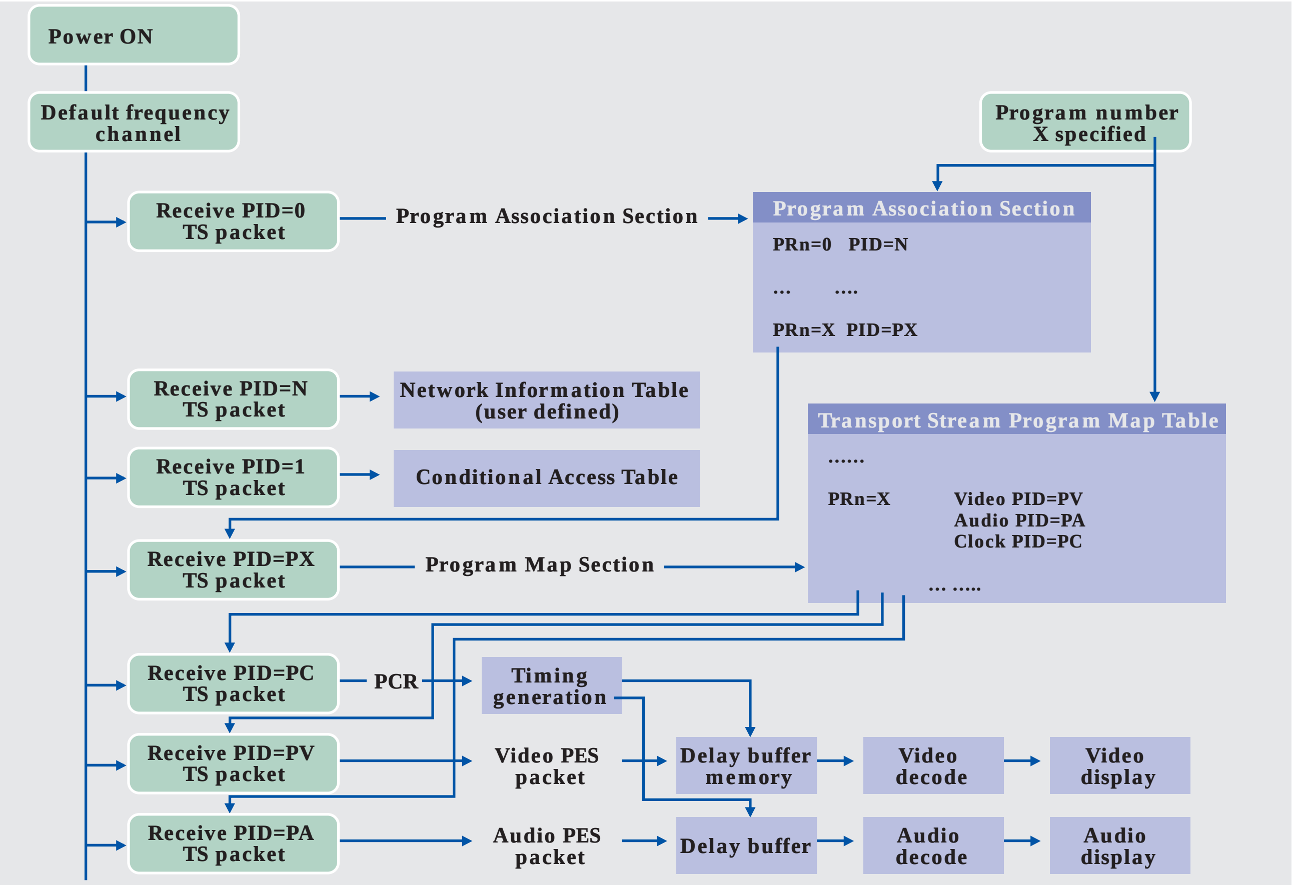


Figure 7: Example of stream selection by PSI in an MPEG2 Transport Stream

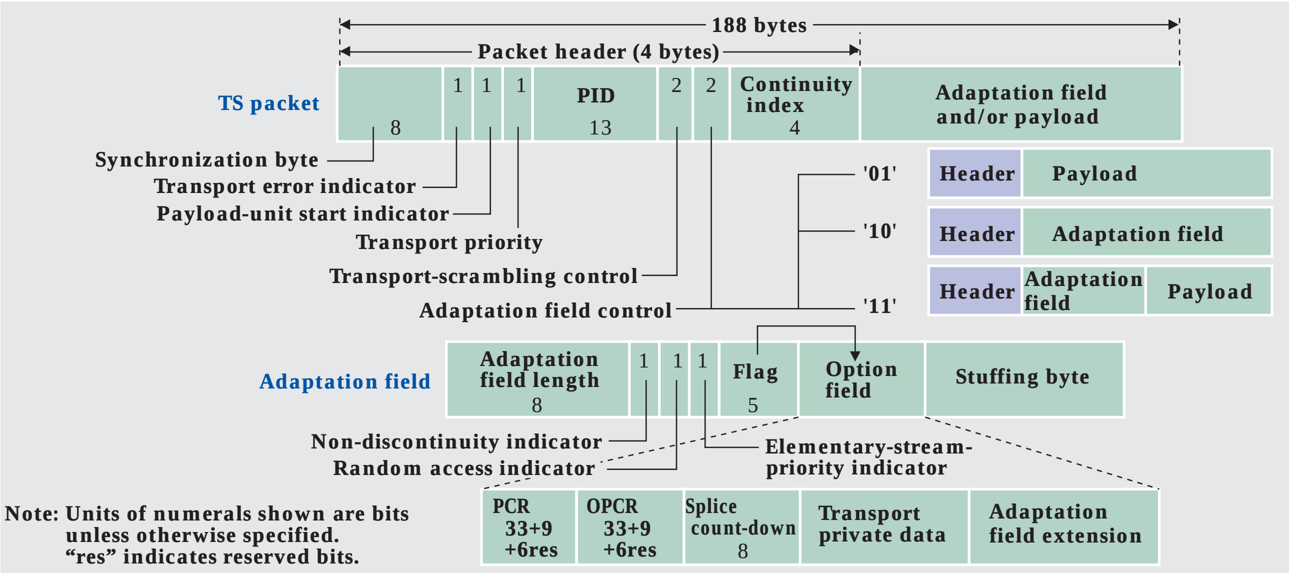


Figure 8: TS packet configuration

stream packets. Each TS packet has a fixed length of 188 bytes. This length is highly compatible with ATM cell lengths and has been chosen because of its applicability to efficient and robust error-correction coding (assuming a maximum coding length of 255 bytes).

The configuration of a TS packet is shown in Fig. 8. A TS packet consists of a packet header (fixed length: 4 bytes) followed by an adaptation field (variable length) as a header extension section and/or payload (data section). An adaptation-field flag (2 bits) in the header indicates the existence of an adaptation field or payload or both following the header.

The most important information in the packet header section is the packet identifier (PID) that can identify about 8000 types of packets in a 13-bit field. TS packets with a PID value of 0, 1, or 2 are special packets that include PSI information. The packet header also includes a payload-unit start indicator (1 bit), transport-scrambling control (2 bits) indicating whether the payload section is scrambled, and a continuity index (4 bits) for detecting packet discontinuity.

One PES packet to be transmitted is divided up among multiple TS packets having the same PID number. The beginning of the PES packet coincides with the beginning of the payload section of a TS packet, which is indicated by setting the payload-unit start indicator bit to '1'.

The adaptation field is for sending various types of information such as splice countdown used to indicate PCR and edit points. It is also used as stuffing (when the final section of a PES packet is placed in a TS-packet payload, dummy data must be added to ensure a fixed-length TS packet).

(2) PES packet configuration

The configuration of a PES packet is shown in Fig. 9. PES packets are used to packetize video and audio data streams

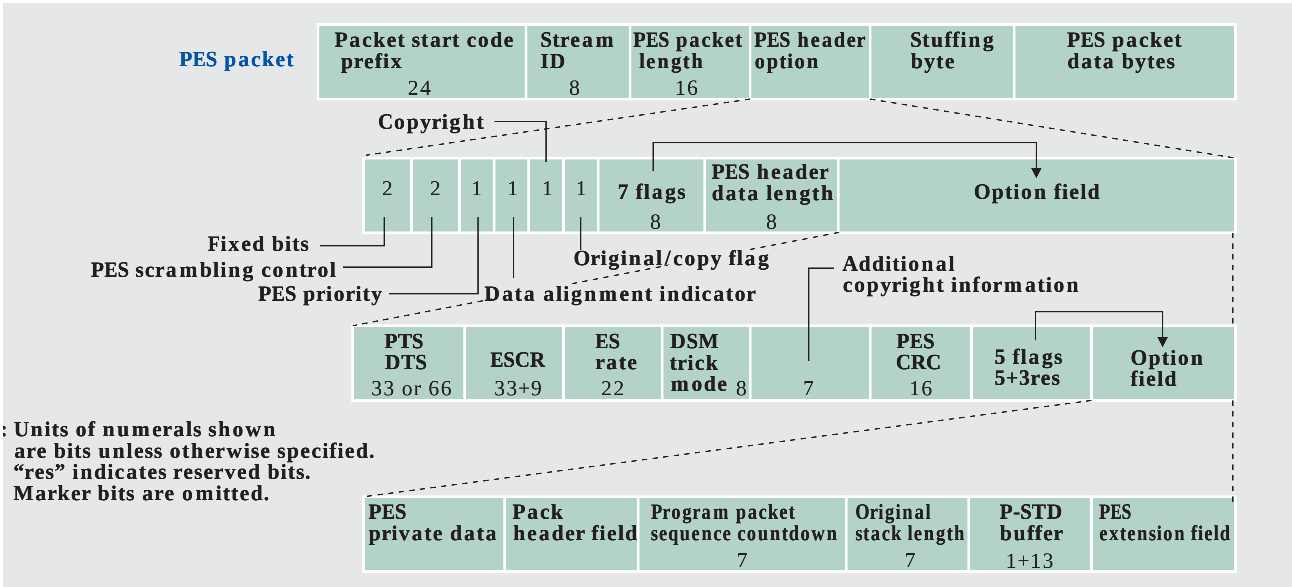


Figure 9: PES packet configuration

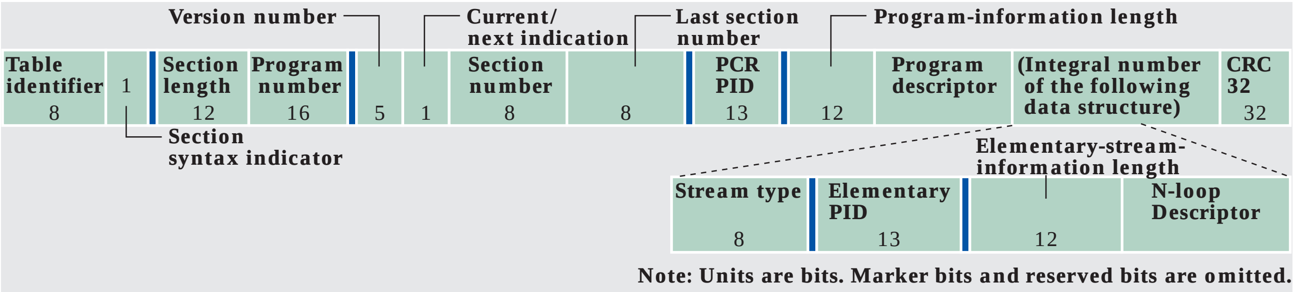


Figure 10: Configuration of TS Program Map Section

Table 5: Summary of MPEG-defined descriptors

Descriptor	Description
Video stream descriptor	Video-encoded version
Audio stream descriptor	Audio-encoded version
Hierarchical descriptor	Hierarchical encoding information for video, audio, and private streams
Registration descriptor	Registration information provided by a registration-control institution to identify private data
Data stream alignment descriptor	Specifies video data alignment (slice, picture, GOP, etc.) or audio data alignment (frame, etc.)
Target background descriptor	Display-format information for background in video window display
Video window descriptor	Window position and overlay priority in video window display
Conditional access descriptor	Information related to conditional access
ISO639 language descriptor	Information related to language of related program elements
System clock descriptor	Information related to system clock used in generating time stamps
Multiplex buffer usage descriptor	Information related to usage policy of T-STD multiplex buffer
Copyright descriptor	Information provided by copyright registration institution to identify copyright
Private data indicator descriptor	Indicates type of private data; descriptor contents are defined by user.
Smoothing buffer descriptor	Information related to buffer operation for smoothing the bit rate in remultiplexing
STD descriptor	Information indicating leak method or VBV delay method in the case of a video stream in T-STD
IBP descriptor	Information describing GOP structure

(and data having similar characteristics) standardized by MPEG.

A PES packet has variable length in units of bytes and consists of a variable-length packet header and variable-length packet data. The packet header includes a stream ID identifying the type of stream in the PS; PTS and DTS time stamps; a control signal for performing scrambling in units of PES packets; a copyright flag; and cyclic redundancy check (CRC) error-detection code.

(3) PSI configuration

A PSI table to be transmitted is placed in the payload of a TS packet in units of "sections." While one TS packet cannot include multiple PES packets, it can contain multiple sections of a PSI table. The beginning of a section may begin in the middle of a payload.

The Program Association Table (PAT) is a one-section table that is transmitted in a TS packet having a PID of '0'. The Conditional Access Table (CAT) is also a one-section table transmitted in a TS packet having a PID of '1'. The Program Map Table (PMT), on the other hand, consists of multiple TS-program map sections that are transmitted in TS packets having PID values arbitrarily specified in the PAT. The configuration of a TS-program map section is shown in Fig. 10. In addition to the above, it is also possible to transmit application-defined sections (referred to as private sections) having the same configuration as that of TS-program map sections.

The section size for PSI tables (PAT, PMT, CAT, TSDT) specified by MPEG is 1K bytes maximum and the size of private sections is 4K bytes maximum.

The start position of a section in a TS packet is indicated by using the payload-unit start indicator in the TS packet header and the pointer in the first byte of the TS packet payload.

A PSI table section includes information such as version number and current-next indication and takes on a structure applicable to partial updating of information. Section data also includes any number of descriptors that provide explanatory information related to an entire program or to individual streams making up a program. Table 5 summarizes the descriptors defined by MPEG. Descriptors having the same basic data structures as these may be specified by individual application systems.

(Mr. Toshiro Yoshimura)

References

(1) ISO/IEC 13818-1:2000, "Information Technology-Generic Coding of Moving pictures and Associated Audio Information-Part1:Systems" (or JIS X4325)

(2) A. MacInnis, "The MPEG systems coding specification", Signal Processing: Image Communications, 4, 153-159 (1992)