

# Development of Media Transport Protocol for 8K Super Hi-Vision Satellite Broadcasting System Using MMT

## ABSTRACT

An ultra-high definition display for 8K Super Hi-Vision is able to present much more information at one time. A variety of services are enabled by presenting various information obtained from broadband networks together with high-quality content delivered through satellite broadcasting channels. In order to realize advanced hybrid services achieved by harmonization of broadband networks with broadcast channels in terms of content delivery, an MMT - based transport technology we have proposed is introduced in an 8K Super Hi-Vision satellite broadcasting system under development in Japan. This report describes the general structure of MMT-based broadcasting systems, carriage of audio/ video signals in the MMT Protocol, and signaling information required for broadcasting systems.

## 1. Introduction

An 8K satellite broadcasting system is being prepared for the start of trial broadcasts in 2016. 8K displays have extremely high resolution, so they are able to present more information than conventional displays. As such, anticipation is growing around new services that present more than just high-quality satellite broadcast content, and include additional related information obtained through broadband networks.

We have been studying how to use such broadband networks with the 8K satellite broadcast system. In

particular, we have proposed a broadcast system using MMT<sup>1)</sup>, a media transport protocol capable of using a variety of broadcast and broadband networks, for the 8K satellite broadcast standard in Japan.

MMT is a standard of the International Organization for Standardization/International Electrotechnical Commission (ISO/IEC). It specifies packet formats, payload formats and control-information formats in order to support a variety of applications, but further study regarding how these specifications are to be used for particular applications is needed. As such, we developed the detailed specifications for implementing a broadcast system and submitted a proposal for a standard to the Association of Radio Industries and Businesses (ARIB). The proposal includes transmission methods for video and audio signals using MMT in a broadcasting system, as well as specifications for various control information. Deliberations at ARIB led to their adopting this proposal as a standard<sup>2)</sup>.

This article describes the standard, proposed by our laboratory and adopted in Japan, including the architecture of the MMT broadcast system, video and audio signal transmission methods, and control information needed to implement the broadcasting system.

## 2. Services Implemented using Hybrid Delivery

Satellite broadcasting transmits stable, high-quality

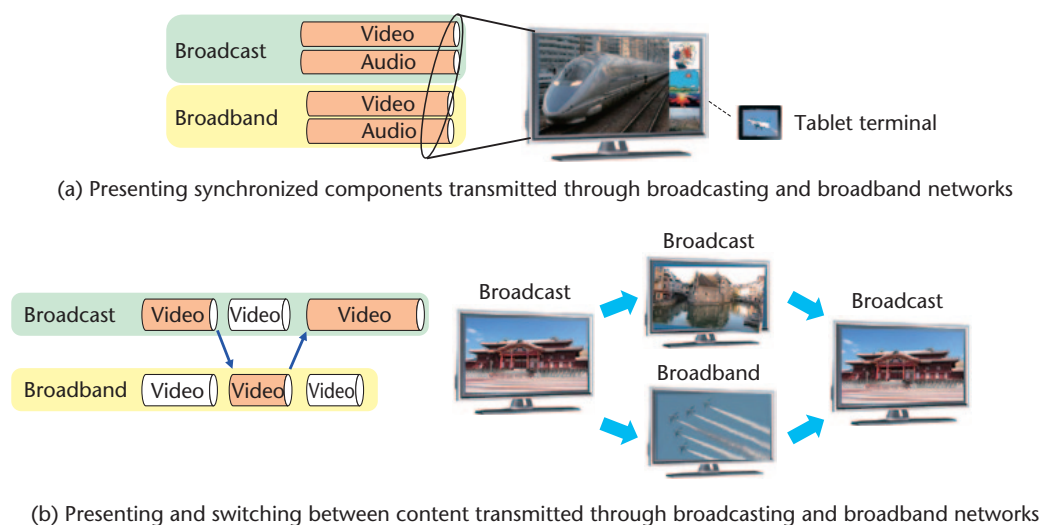


Figure 1: Examples of services made possible with hybrid delivery

content to many users simultaneously. Conversely, broadband network is capable of two-way transmission and can transmit content at the request of the receiver. Here, a hybrid form of delivery using both satellite broadcasting and broadband network would be able to provide services such as those shown in Figure 1.

Figure 1(a) is an example of a service that presents synchronized video, audio and other components (elements that comprise the content) that have been transmitted by broadcasting and broadband. Components that are expected to be viewed by most of the users are transmitted by broadcasting, while individually requested components are provided through broadband. In doing so, video components are presented at a specified position on the screen, so that the ultra-high-definition display can be used most effectively. Content related to the broadcast can also be viewed on tablet terminals, which are becoming more and more popular.

Figure 1(b) is an example of a service that tailors the presented content, such as by changing the commercials displayed according to the age or sex of the viewers. Content especially suited to the individual viewers can be presented by switching between audio and video received through broadband channels or saved earlier.

By delivering content with this sort of hybrid delivery, we expect new services will arise that use the television

in ways not possible before<sup>3)-6)</sup>.

## 3. Overview of MMT Broadcasting System

### 3.1 Broadcast System Architecture

MMT is a media transport protocol suitable for delivering content using a variety of transmission paths, including one-way transmission paths such as broadcasting<sup>7)8)</sup>. It was approved as an international standard by the ISO/IEC JTC1/SC29/WG11 Moving Picture Experts Group (MPEG) in March, 2014.

Our proposal, which is based on MMT, was published as part of the March, 2014, report from the Information and Communications Council in Japan. It is in the section on “Technical requirements for backbone and public satellite broadcasting” within “Technical requirements for ultra-high-definition television broadcasting systems,” for implementing advanced services linking broadcasting and broadband networks and capable of using both types of transmission paths. This report formed the basis of the ARIB standards for implementing the 8K satellite broadcasting system.

The architecture of the 4K/8K broadcasting system specified in the ARIB standard is shown in Figure 2. In the layer model of broadcast transmission path in Figure 2 (a), encoded video and audio signals are encapsulated in MMT protocol (MMTP) packets and

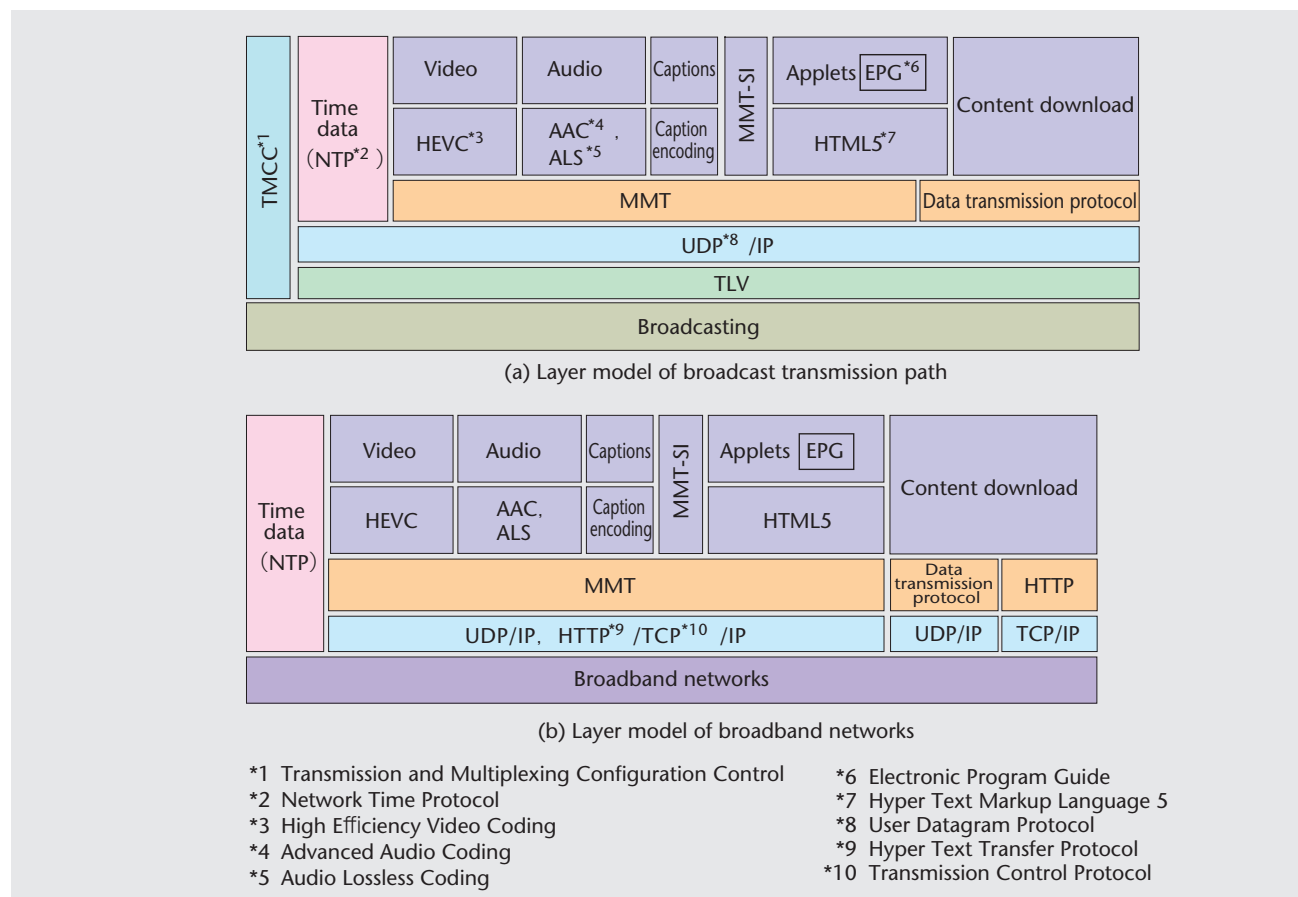


Figure 2: 4K/8K Broadcasting system architecture

transmitted inside Internet Protocol (IP) packets. The Type-Length-Value multiplexing<sup>9)\*1</sup> is used for efficient multiplexing of IP packets. In addition to transmitting video and audio signals in this way, control information for broadcast programs is defined as the MMT Signaling Information (MMT-SI). The layer model of broadband networks in Figure 2(b) is very similar to the layer model of broadcasting transmission path in Figure 2(a), thanks to the feature of MMT that broadcasting and broadband transmission paths can be handled in the same way.

We have implemented this broadcast system and demonstrated content delivery in which different content components are transmitted on broadcast and broadband networks and synchronized and presented on the receiver equipment. These demonstrations showed the effectiveness of MMT, and the system was exhibited at our laboratory open house in 2014.

### 3.2 Relation between MMT Packages and Broadcast Services

A unit of content in MMT is called a package. The relationship between these packages and a broadcasting service is shown in Figure 3. A broadcasting service is a succession of programs sent according to a schedule.

In the MMT-based broadcasting system, packages and services are associated in a one-to-one relationship. In conventional broadcasting systems, broadcast services had a one-to-one association with programs, which are specified in MPEG-2 Systems<sup>2</sup> (the unit of content for MPEG-2 Systems), but the MMT broadcast system uses packages instead of programs. For a given service (package), a program delineated by a start time and an end time is called an event.

Multiple packages can be multiplexed and transmitted

on a single IP data flow, as shown in Figure 3. Here, an IP data flow is the set of all IP packets for which five fields included in IP and User Datagram Protocol (UDP) headers; the source IP address, the destination address, the protocol type in the IP header, the source port number (specifying the sending application), and the destination port number (specifying the receiving application); have the same values. In addition to IP data flows for package transmission, broadcast transmission paths also have IP data flows for download services and extension services.

These multiple IP data flows are multiplexed into Type-Length-Value (TLV) streams, which are the transmission units for the broadcast transmission path. Here, by stream, we mean “flow of data,” regardless of the type of transmission path or signal format. A TLV stream is a sequence of TLV packets identified by a TLV stream identifier (ID) and a TLV packet with control information for the TLV multiplexing format (TVL-SI), including a Network Information Table (TVL-NIT) and an Address Map Table (AMT). TLV-NIT is control information for the transmission path, such as the broadcast signal frequencies. The AMT is control information indicating the multicast group<sup>3</sup> list for packets comprising the broadcast service. The transmission slot containing these TLV packets is identified by the TLV stream ID in the Transmission and Multiplexing Configuration Control (TMCC) signal of the satellite broadcast.

The service architecture when using both broadcast and broadband networks is shown in Figure 4. In the figure, video component 1, audio component 1, and data 1 are transmitted on the broadcast transmission path, and video component 2, audio component 2, and data 2 are transmitted on the broadband path. The three components transmitted on the broadcast transmission path are multiplexed into the same IP data flow and transmitted in the same TLV stream. This is because

\*1 A framework for multiplexing IP and other variable-length packets.

\*2 A media transport protocol standardized in 1994 by MPEG and currently used for digital broadcasting.

\*3 The destination IP address used in the case of IP multicast delivery.

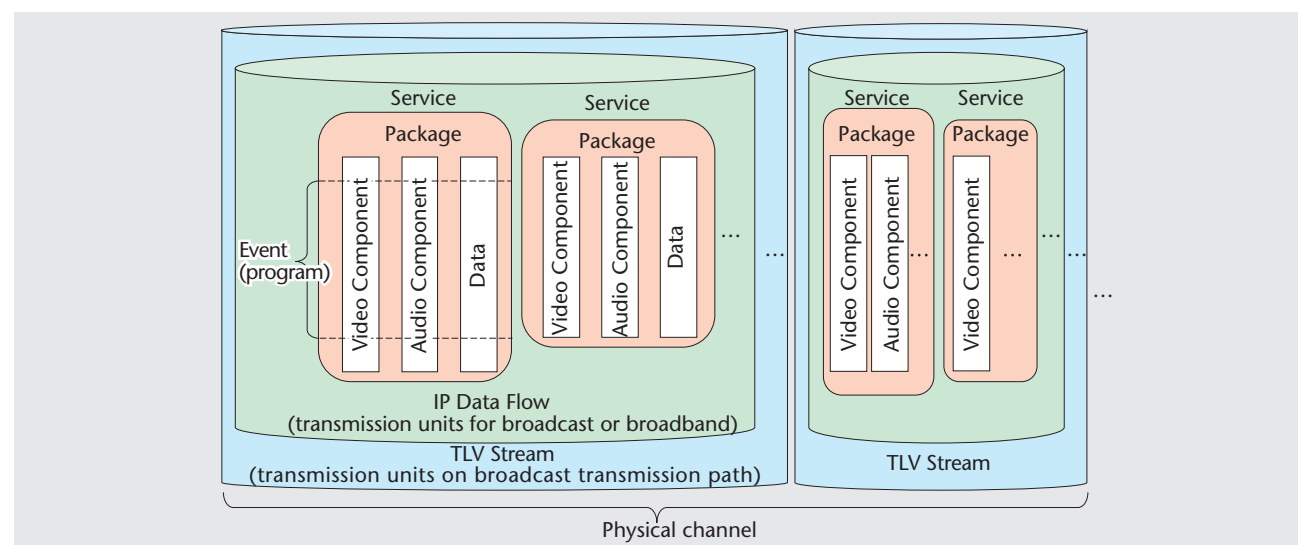
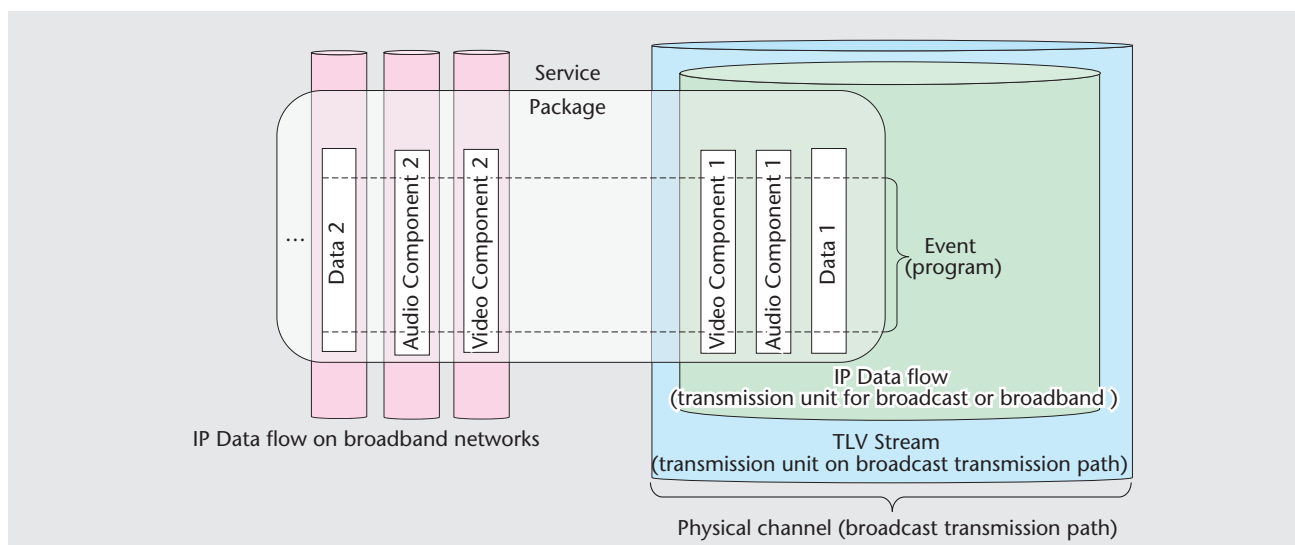


Figure 3: Relation between MMT Packages and broadcast services



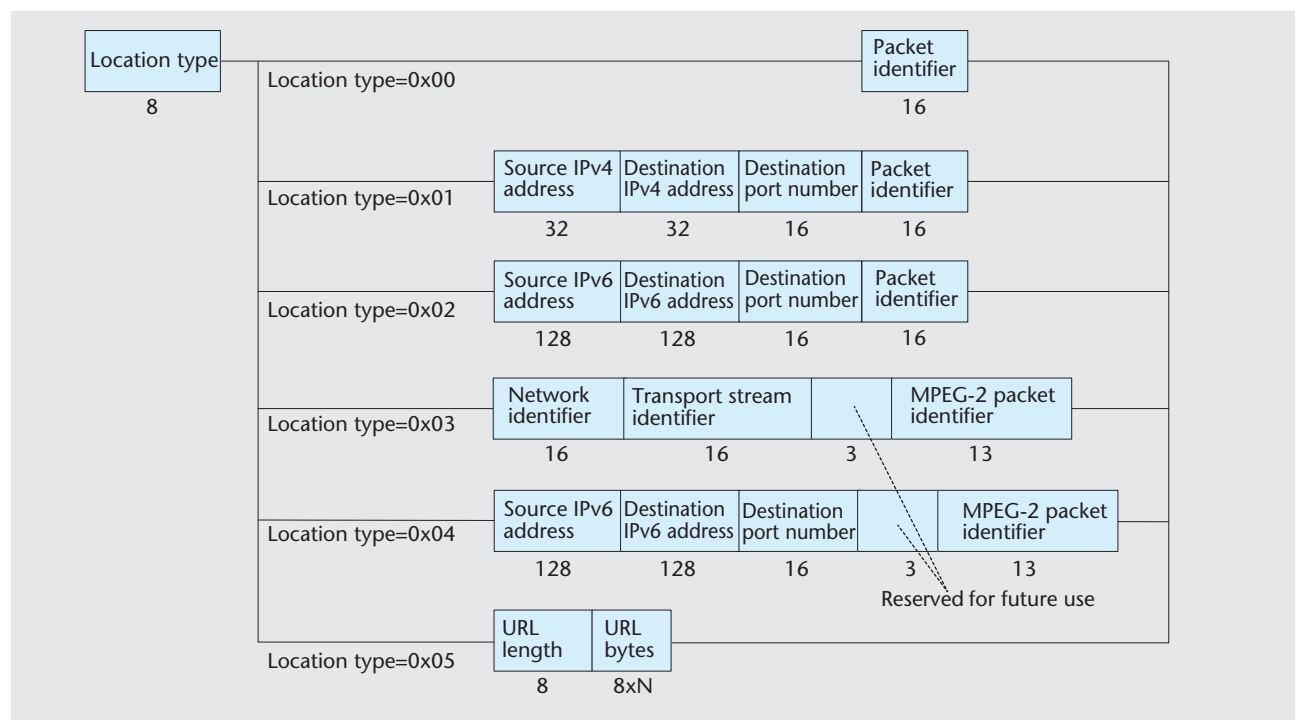
**Figure 4: Service architecture when using both broadcast and broadband transmission**

all information transmitted on the broadcast path is transmitted to the receiver, and separate IP data flows are not needed. On the other hand, the components transmitted on the broadband networks are transmitted in response to individual requests from the receivers, so they are transmitted using IP data flows for each component.

MMT has control information called the MMT Package (MP) table. The MP table has details such as the types of components that comprise packages and where to get them.

Where to get components is specified in the location information (general\_location\_info) in the MP table

structure shown in Figure 5. The location information is one of six location types depending on how the component source is specified. It specifies the locations of components multiplexed in the same IP data flow as the MP table (location type=0x00), in the MMTP packets on arbitrary IP data flows (location type=0x01 or 0x02), on broadcast networks using MPEG-2 Transport Streams (TS) (location type=0x03) or in MPEG-2 TS packets transmitted with MPEG-2 TS over IP (location type=0x04). Locations of components can also be specified using a Uniform Resource Locator (URL) (location type=0x05). In particular, the locations of components comprising the content can span a number of broadcasting and



**Figure 5: Component source based on location information (numbers represent the number of bits in each field)**

broadband networks sources.

In this way, components transmitted over different transmission paths with MMT can be gathered into one package, making it easy to implement services using both broadcasting and broadband networks.

### 3.3 MMT Signaling Units

Encoded signaling units used in MMT include Media Processing Units (MPU), Media Fragment Units (MFU), the MMTP payload, and MMTP packets.

The MPU is the unit of data processing. The general structure of an MPU is shown in Figure 6. An MPU is composed of MPU metadata, which describes the overall attributes of the MPU, movie fragment metadata, which describes attributes such as access units<sup>4</sup>, and sample data, which includes the video and audio signals. An MPU can include one or more access units, and a single MPU is a unit for which video and audio can be decoded. For video signals using inter-frame predictive coding, an MPU must be the same unit as a Group of Pictures (GOP). An MPU has a sequence number for MPUs in the same component. MPUs can be differentiated from each other by a combination of the asset ID, which identifies the component, and the per-MPU sequence number.

MFUs are smaller than MPUs and are used when

partitioning sample data. For video signals, the Network Abstraction Layer (NAL) unit<sup>5</sup> is the MFU, and for audio signals, the access unit is the MFU. Thus, by composing transmissions of MFUs, which are smaller than MPUs, transmission errors in the video and audio signals decoded at the receiver can be minimized, even if there is packet loss or other decreases in transmission quality.

MFUs and control information are transmitted using MMTP packets. MMTP packets are composed of a header and a payload. When the MFU or control information to be transmitted is small compared with the Maximum Transmission Unit (MTU), which is determined by the transmission path, multiple MFUs of the same type or multiple control data items can be encapsulated together in one MMTP payload. On the other hand, if the MFUs or control information being transmitted are larger than an MTU, they can be fragmented and encapsulated in multiple MMTP payloads (Figure 7).

One MMTP payload is sent as a single MMTP packet. A single MMTP packet cannot send multiple MMTP payloads, and neither can one MMTP payload span multiple MMTP packets.

There are two ways that an MMTP payload can be

<sup>4</sup> The smallest synchronizable unit, such as a video frame or audio encoding unit.

<sup>5</sup> A unit of data, such as video signal motion vectors, arithmetically coded orthogonal transform coefficients, and supplementary data.

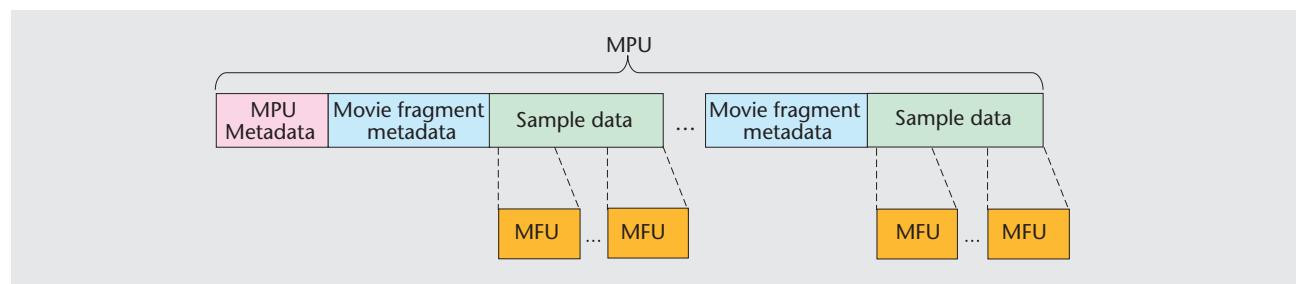


Figure 6: General structure of MPU

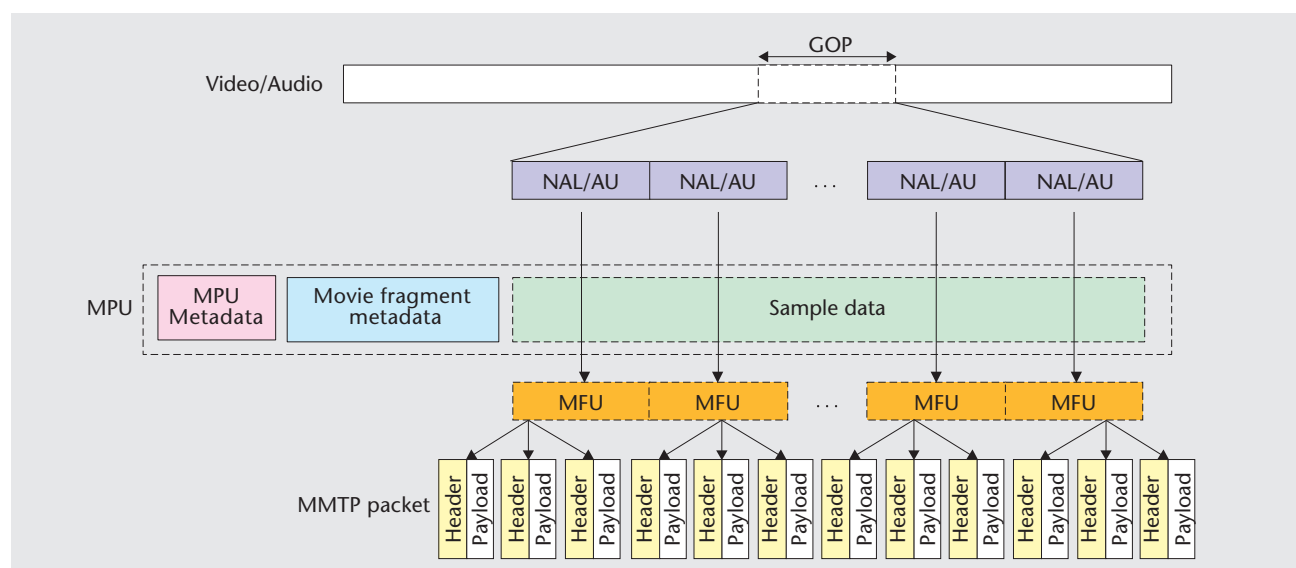


Figure 7: Example of one MFU fragmented into multiple MMTP payloads

composed from MPUs and MFUs. One way is to compose an ordinary MPU from the video signal output by the encoder, as shown in Figure 6, and to fragment the MPU into MMTP payloads. The other way is to omit the process of composing ordinary MPUs as shown in Figure 6, and to take the video and audio signal NAL units and access units output from the encoders as MFUs and encapsulate them in the MMTP payload. The latter can be used in broadcasting to reduce transmission delays. In this case, the additional information needed for the MPU is sent in the MMTP packet header.

#### 4. Video and Audio Signal Transmission in the MMT Broadcast System

##### 4.1 Video Signal Transmission using MMTP

When transmitting video signals encoded with High Efficiency Video Coding (HEVC), the encoded video signal is input to the MMT processing in the form of NAL units. NAL units are classified as either Video Coding Layer (VCL) NAL units, which hold arithmetically encoded video signal motion vectors or orthogonal transform coefficient data, or non-VCL NAL units, which hold supplementary data for VCL NAL units. Examples of NAL units are shown in Table 1. With MMT, all NAL units are handled as MFUs.

An overview of how NAL units comprising an MPU are turned into MMTP packets is shown in Figure 8. As

mentioned earlier, video signal MPUs must be the same as GOP units, so an Intra Random Access Point (IRAP)<sup>\*6</sup> access unit is needed at the beginning of an MPU. Thus, in Figure 8, a CRA NAL unit, which comprises the IRAP access unit, is placed before the MPU VCL NAL units. Other non-VCL NAL units needed to decode the video signal are also placed before this CRA.

The size of an MPU affects the amount of delay between switching broadcast channels and when the video is displayed. If the MPU comprised multiple GOPs, the delay would increase. For this reason, one video signal MPU is composed of a single GOP.

##### 4.2 Supporting Hierarchical Encoding in the Time Direction for Video Signals

It is desirable that 120 frame per second video signals (120P<sup>\*7</sup>) also be able to be decoded at 60 frames per second (60P) and displayed as 60P video. To support this functionality, the signal structure must enable the 60P sub-bit stream to be separated from the 120P subset of the HEVC stream, which is hierarchically encoded in the time direction, and these must be transmitted such that receivers can distinguish between the two. The sub-bit stream contains the information needed to decode the

<sup>\*6</sup> An access unit that can be decoded immediately.

<sup>\*7</sup> A progressive scan video signal with a 120 Hz frame frequency.

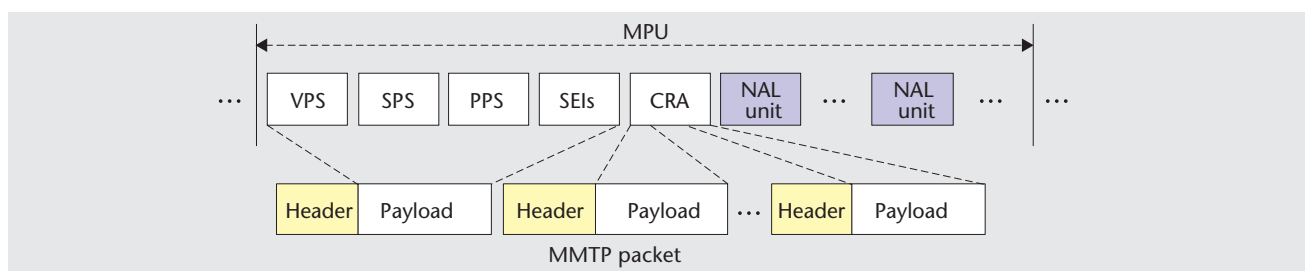
**Table 1: Examples of NAL units in the HEVC standard**

(1) Non-VCL NAL units

| NAL unit name                                     | Description  |
|---|--|
| VPS (Video Parameter Set)                         | Gives shared and relationship information between layers in hierarchical encoding. |
| SPS (Sequence Parameter Set)                      | Gives common parameters for video sequences.                                       |
| PPS (Picture Parameter Set)                       | Gives common parameters for video pictures.  |
| Prefix-SEI (Supplemental Enhancement Information) | Supplemental enhancement information placed in front of a VCL NAL unit.            |

(2) VCL NAL unit

| NAL unit name                        | Description  |
|--------------------------------------|--|
| CRA (Clean Random Access)            | An IRAP containing data that enables decoding of video frames following this one in the display order. |
| IDR (Instantaneous Decoding Refresh) | An IRAP containing data that enables decoding of video frames after this one in the display order.     |
| TRAIL                                | Decoded after an IRAP video frame, including data displayed before the IRAP video frame.               |



**Figure 8: Summary of NAL units and MMTP packetization comprising an MPU**



video at a specific time resolution. In this section, we discuss this signal structure and how it is transmitted.

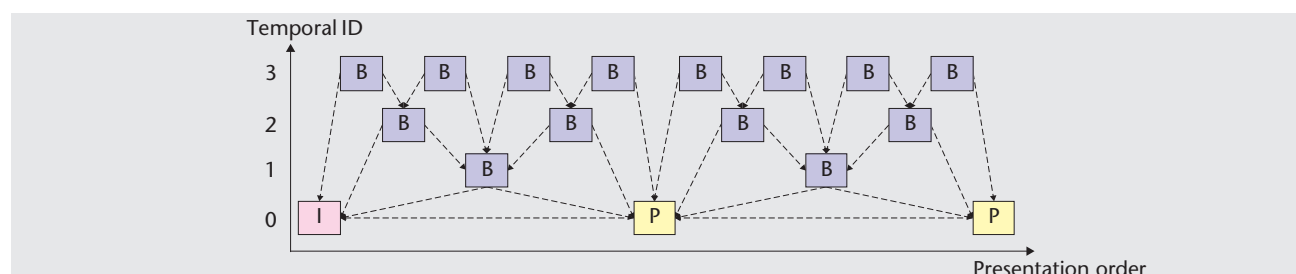
Figure 9 shows an example of the structure of an HEVC stream hierarchically encoded in the time direction. Here, I represents video frames for which inter-frame prediction is not done, P represents frames for which prediction in the forward direction is done, and B represents frames with prediction done in both directions. NAL units with a Temporal ID (a video frame time-direction identifier) of 0 to 2 comprise the 60P sub-bit stream, and those with a Temporal ID of 3 comprise the 120P subset. Figure 10 is an overview of how the time-direction hierarchically encoded HEVC bit stream is transmitted. When the package is composed, the 60P sub-bit stream and 120P subset are separated into different components. In the figure, the former is shown as Asset (a video or audio component) 1 and the latter as Asset 2. Since Asset 1

and Asset 2 are different components, the access units that they contain are transmitted in MMTP packets with different packet IDs.

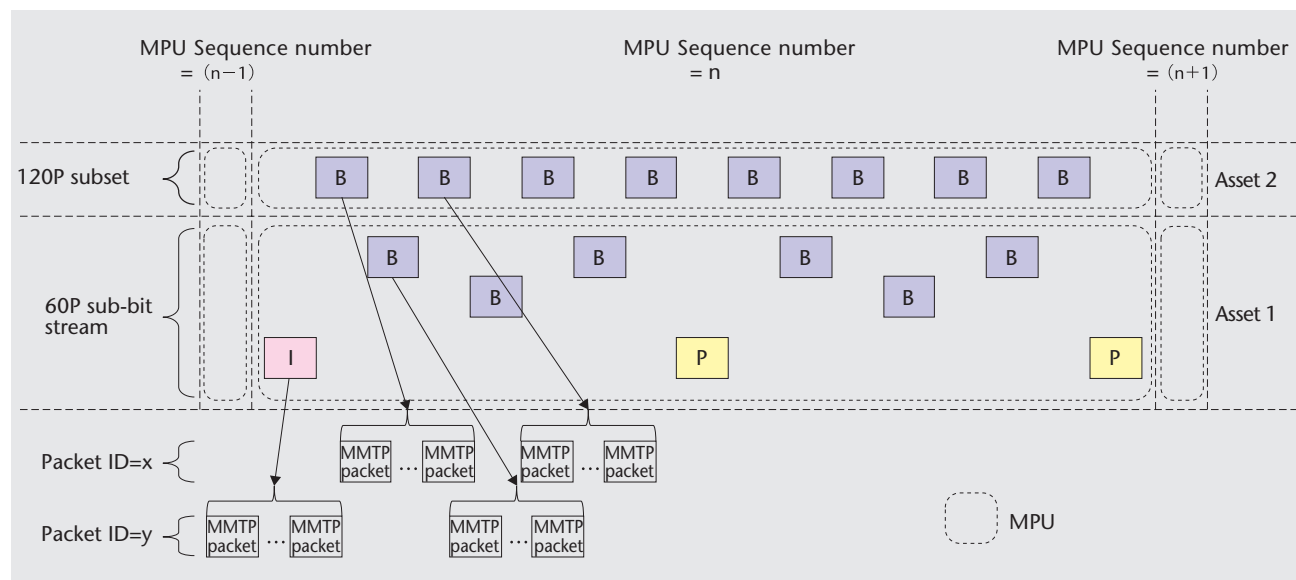
Only Asset 2 cannot be decoded. Asset 1 is needed to decode Asset 2, so a dependency descriptor (see Table 4) is inserted in its MP table in the descriptor area to describe Asset 2 should be decoded with Asset 1 together.

The same MPU sequence number as Asset 1 is also added to Asset 2 so that the temporal relationship between the assets can be easily determined. Thus, by adding the same MPU sequence numbers to MPUs that are related in time, receivers can easily identify which MPUs in Asset 1 are needed to decode MPUs in Asset 2.

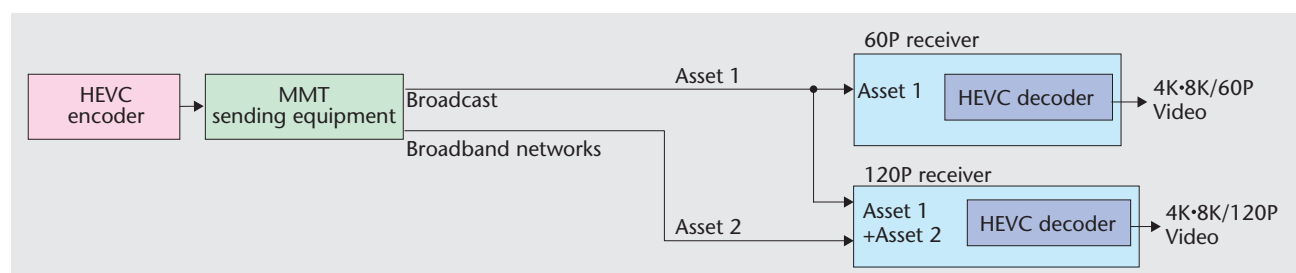
In the manner described above, receivers supporting 120P can use a 120P HEVC decoder to decode and display 120P video, and those supporting 60P can use a 60P HEVC decoder to decode and display 60P video.



**Figure 9: Example of hierarchical encoding in the time direction**



**Figure 10: Overview of transmission of time-direction hierarchical encoding HEVC bit stream with MMT**



**Figure 11: Supporting 120P with hybrid delivery**

Also, in addition to broadcasting 120P video, it is also possible to broadcast the sub-bit stream for 60P, while sending the 120P subset through a broadband network. As shown in Figure 4, an MMT package can include both components sent by broadcasting and others sent by broadband networks. Thus, Asset 1 can be sent through the broadcasting channel, while Asset 2 is sent through the broadband networks, and receivers can receive Asset 2 and support 120P as needed (Figure 11).

#### 4.3 Audio Signal Transmission using MMT

It is expected that the next generation of broadcasting systems will use MPEG-4 Advanced Audio Coding (AAC) or MPEG-4 Audio Lossless Coding (ALS). These encoding schemes use the Low Overhead MPEG-4 Audio Transport Multiplex (LATM)/Low Overhead Audio Stream (LOAS) stream formats. LATM provides information regarding the audio data channel configuration and multiplexing, such as order and concatenation. LOAS provides a synchronization function. The LATM/LOAS stream format is a sequence of AudioMuxElements<sup>8</sup>, including one or more audio frames, and is specified in the MPEG-4 Audio standard<sup>10</sup>.

When transmitting the audio signal in this way, the encoded audio signal can be processed with MFUs being the minimal units by taking a single AudioMuxElement to be one MFU. Unlike the video signals, MPEG-4 AAC and ALS do not encode over multiple audio frames, so any audio frame can be used as a random access

point<sup>9</sup>. This also allows MPUs to be composed of a single audio frame. However, in that case, an MP table must be sent for each audio frame to specify the presentation time for the audio signal, which increases overhead<sup>10</sup>. Alternatively, an MPU could be composed of multiple audio frames. To simplify the description of the MP table, which references both video and audio signals, it would be suitable to compose MPUs of audio frames equivalent to the IRAP interval of the video signal.

#### 5. MMT Broadcast System Control Information

##### 5.1 Types and Function of Control Information

Control information is signal expression information related to the organization of broadcast programs and services, and there are three types: messages, tables, and descriptors. Messages are control information for storing tables and descriptors when they are sent. Tables are control information describing elements and attributes that express specific information. Descriptors express more detailed information. The main messages, tables and descriptors, together with their functions, are given in Tables 2, 3, and 4.

##### 5.2 Control Information for Presentation

When presenting multiple videos on an ultra-high-definition 8K display with 7,680×4,320 pixels, the

<sup>8</sup> A bit stream including information such as audio data or decoder initialization data.

<sup>9</sup> A starting position in data from which data can be decoded correctly.

<sup>10</sup> Additional data besides the data itself, such as headers and tables.

**Table 2: Main messages**

| Message name                    | Function   |
|---------------------------------|--|
| PA (Package Access) message     | A package that can include all tables. Control information that the receiver processes first.                                    |
| M2 Section message              | Transmits MPEG-2 system section extension formats for transmitting current digital broadcasting program arrangement information. |
| CA (Conditional Access) message | Transmits information related to conditional access system.  |

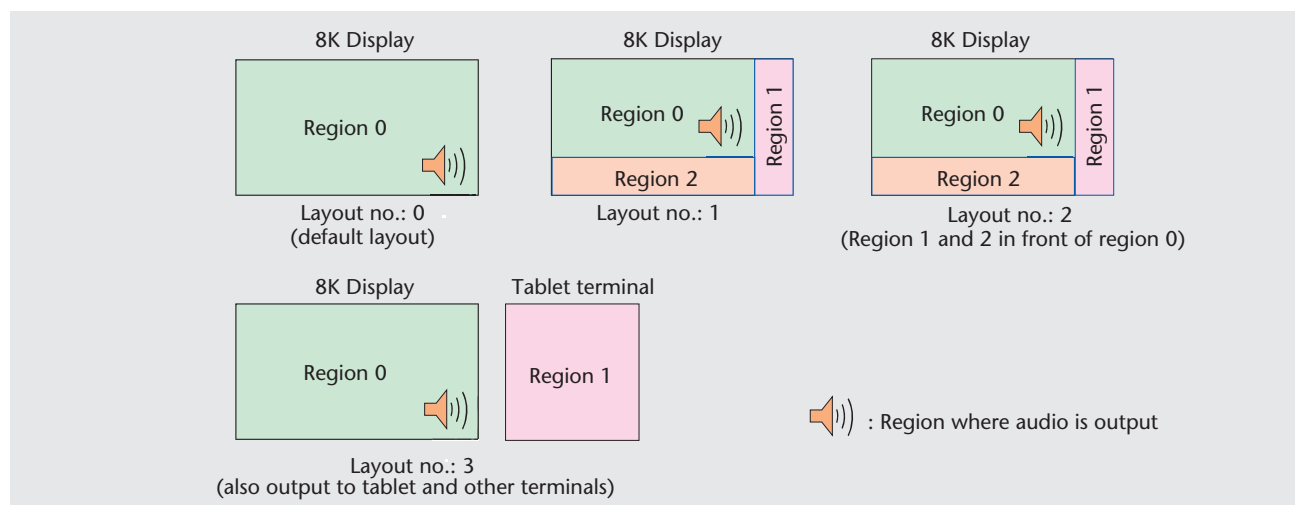
**Table 3: Main tables**

| Table name                 | Function  |
|----------------------------|---|
| MP (MMT Package) table     | Gives information for composing a package, such as component types and where to get them.   |
| Package list table         | Gives list of IP data flows transmitting PA messages as well as packet IDs and IP services. |
| Layout configuration table | Associates layout numbers with presentation layout information.                             |
| CA table                   | Transmits identifiers for the conditional access system.                                    |

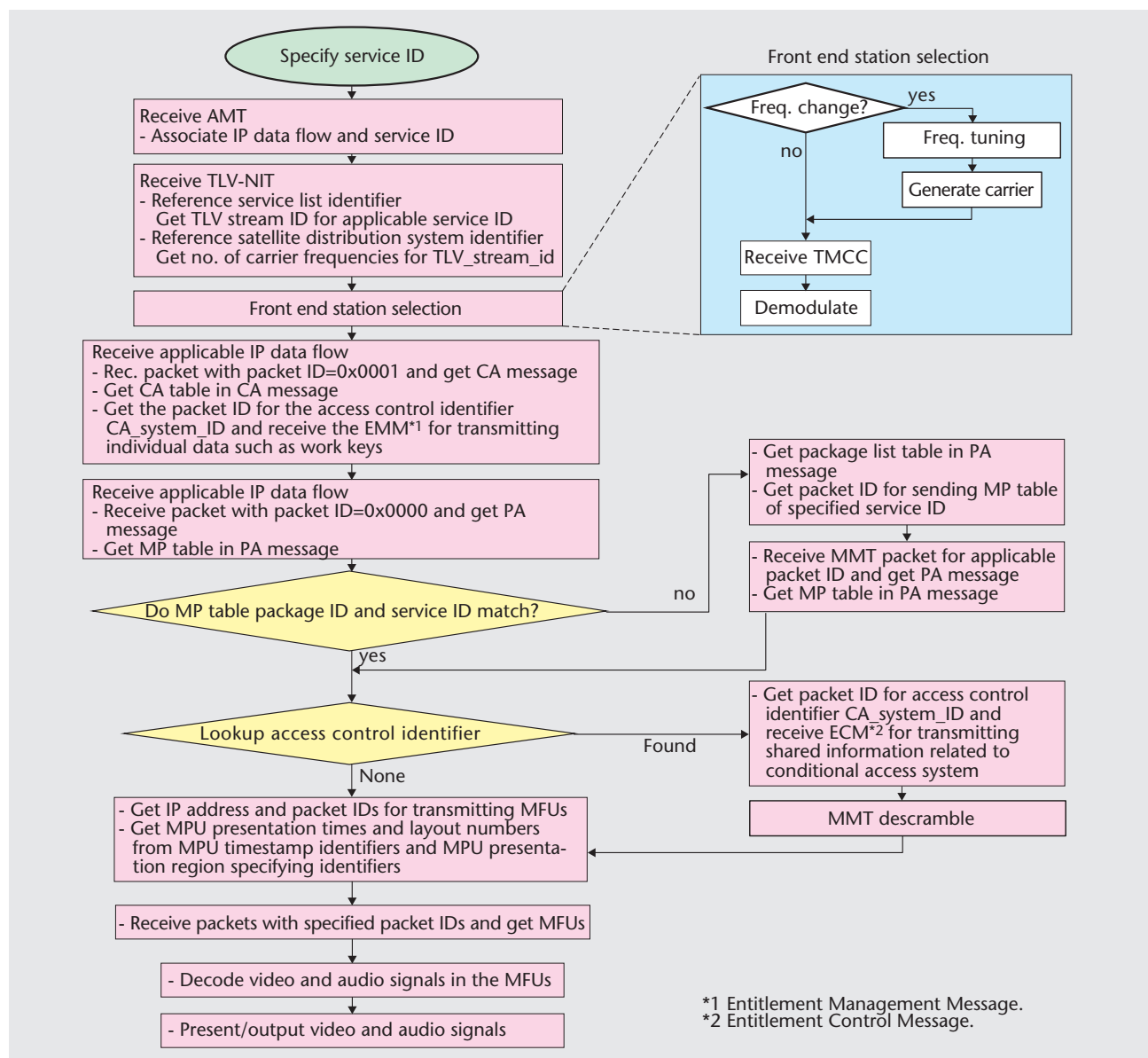
**Table 4: Main descriptors**

| Descriptor name                    | Function   |
|------------------------------------|--|
| Asset group descriptor             | Gives asset group relation and precedence within group.                    |
| MPU presentation region descriptor | Gives MPU presentation location.   |
| MPU timestamp descriptor           | Gives MPU presentation time.   |
| Dependency descriptor              | Gives asset IDs of components with dependency relationships.               |
| Emergency information descriptor   | Describes required information and function for emergency warning signals. |





**Figure 12: Examples of layouts specifying video presentation regions**



**Figure 13: Control information processing when selecting channels on a receiver**

information can be displayed effectively by specifying the display position for each of the videos. To this end, the MMT control information allows for the presentation location to be specified by specifying a layout number and region number in addition to the presentation time for each video (Figure 12). Specifically, region allocations are done using a layout settings table, which is one type of control information. This type of control information makes it possible for services to use ultra-high-definition displays.

### 5.3 Processing Control Information when Selecting Channels

Figure 13 shows how the receiver uses control information when selecting channels.

Channels are selected by specifying the desired service ID. When a user specifies a service ID, the receiver uses an AMT to associate the service ID desired by the user with the IP data flow information. Then, the physical channel for the corresponding service ID is identified using a TLV-NIT, and this physical channel is selected. As a result, the desired IP data flow is output from the front end of the receiver.

The output IP packets contain MMTP packets. From them, the MMTP packet with a value of 0x0000 in the ID field of the MMTP packet header is selected to get the Package Access (PA) message. The PA message can contain all of the tables, as indicated in Table 2. The receiver gets the MP table from inside the PA message.

Depending on the broadcast time, a single IP data flow might contain multiple packages (Figure 3). Thus, the package ID of the retrieved MP table is checked to ensure it matches the service ID desired by the user. If it does not, the package list table in the PA message is retrieved to identify the packet ID of the MMTP packet transmitting the MP table of the required service ID.

The location information in the MP table is used to identify the IP data flows and IDs of the packets transmitting the components that comprise the desired content. At the same time, the MPU presentation time and layout number are determined from the MPU timestamp descriptor and MPU presentation region descriptor.

Finally, the MMTP packets transmitting the components are received and the necessary MFUs are obtained. The video and audio signals in these MFUs are decoded and presented or output at the specified presentation time and location from the layout and region numbers.

## 6. Conclusion

In this article, we have discussed the organization of the MMT broadcast system, how the video and audio signals are transmitted, and the variety of control information especially as it relates to standardization of 8K satellite broadcasting.

The MMT broadcast system has been standardized by ARIB, and as of March, 2015, studies on technical reports for practical operation are ongoing. We intend to continue development and testing of transmission methods using broadband networks in order to create new services that make use of both broadcasting and broadband, as only possible with MMT.

This article was revised and amended based on the following papers appearing in IEEE Transactions on Broadcasting.

Y. Lim, S. Aoki, I. Bouazizi and J. Song: "New MPEG Transport Standard for Next Generation Hybrid Broadcasting System with IP," IEEE Transactions on Broadcasting, Vol. 60, No. 2, pp. 160-169 (2014)

(Shuichi Aoki)

## References

- 1) ISO/IEC 23008-1:2014, "Information Technology - High Efficiency Coding and Media Delivery in Heterogeneous Environments - Part 1: MPEG Media Transport (MMT)"
- 2) ARIB: "MMT-Based Media Transport Scheme in Digital Broadcast Systems," ARIB STD-B60 (2014)
- 3) H.-J. Lee, P. Kim, T. Kim and D.-G. Oh: "Broadband Systems Based on DVB-S2 and Mobile DVBRCS, and Their Future Applications to Broadband Mobiles," International Workshop on Satellite and Space Communications, pp. 98-102 (2006)
- 4) B. D. Lee: "Provisioning of Rich Media Experiences for Hybrid Mobile Broadcast Services," IEEE Transactions on Consumer Electronics, Vol. 55, No. 3, pp. 1113-1120 (2009)
- 5) P. Siebert: "Hybrid Broadband/Broadcast Systems and Set-top Boxes," IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (2008)
- 6) A. Bria: "Future Hybrid Cellular-Broadcasting Systems for Multimedia Multicasting," International Conference on Wireless and Mobile Communications (2006)
- 7) Y. Lim: "IP-Friendly MPEG Media Delivery Standards for Next Generation Broadcasting," IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (2012)
- 8) Y. Lim, K. Park, J. Lee, S. Aoki and G. Fernando: "MMT: An Emerging MPEG Standard for Multimedia Delivery over the Internet," IEEE Multimedia Mag., Vol. 20, No. 1, pp. 80-85 (2013)
- 9) Rec. ITU-R BT.1869, "Multiplexing Scheme for Variable-length Packets in Digital Broadcasting Systems" (2010)
- 10) ISO/IEC 14496-3:2009, "Information Technology - Coding of Audio-visual Objects - Part 3: Audio"