Design and Implementation of 8K UHD Encapsulation Method for Efficient Transmission and Reception based on MMT

Seulki Song, Youngsu Ryu, Jungwook Wee, Kyungwon Park, and Kiwon Kwon
Korea Electronics Technology Institute, Seongnam-si, 463-816 Republic of Korea
[e-mail: {seulkiemma , fernwehist, jwwee, kwpark, kwonkw} @keti.re.kr]
*Corresponding author: Kiwon Kwon

Abstract

In this Paper, we propose 8K UHD (Ultra High Definition) encapsulation method for efficient transmission and reception based on MMT (MPEG Media Transport). Broadcasting services for 8K UHD allow users to feel the maximized reality. However, present technology is difficult to provide 8K UHD in broadcasting networks, because the 8K UHD bitrate is too high to be transmitted in the current broadcasting networks. Research for transmitting 8K UHD is underway. In some researches, a receiver is implemented with four 4K UHD display instead of a 8K UHD display. In order to transmit 8K UHD within the limited transmission bitrate of broadcasting network, 8K UHD contents encoded by SHVC (Scalable High Efficiency Video Coding) and then transmitted over heterogeneous network. For using the broadcasting and communication networks, MMT standard is used. MMT is IP based transmission protocol as the next generation transmission protocol. According to the MMT standard, video stream encapsulated and transmitted in MMTP (MMT Protocol) packet. IP-based broadcasting and communication networks can be used to transmit simultaneously, and the receiver can synchronize and play it. We propose an encapsulation method that can efficiently transmit and receive 8K UHD. The proposed method increases a payload rate and decreases an initial delay at the receiver. We show that the efficiency of the proposed method is verified by experimental tests.

1. Introduction

Recently, development of broadcasting and communication networks is evolving at a rapid pace. Users demand high-quality and high-difinition broadcast services. 8K UHD is a broadcasting service that allows users to feel the maximized reality [1]. To transmit 8K UHD contents, however, many technologies such as content acquiring, video coding and transmission are required. A bitrate of 8K UHD is four times higer than the 4K UHD at least shown in **Table 1**. The 8K UHD bitrate is too high to be transmitted in the current broadcasting networks**.**

Table 1. Comparison of the video resolution

Format	Size	Color Depth	Frame Rate	Stream Rate (Mbps)
Full HD	1920x1080	24	30	4~5
4K UHD	3840x2160	36	30	22~28
8K UHD	7680x4320	36	60	88~112

The 8K UHD TV is also not launched yet. For 8K UHD broadcasting service, many studies are underway. Sohn, Y. et al studied 8K UHD broadcasting services [2]. They proposed that the system divides a 8K UHD content into four 4K parts, and then delivers them over heterogeneous. In order to transmit 8K UHD contents within the limited transmission bitrate of broadcasting network, 8K UHD contents are encoded by SHVC for separate transmission over the heterogeneous networks [3]. Then, it transmits the lower quality layer depending on the limited transmission bitrate of the broadcasting network and transmits high quality layer such as 8K UHD through the communication network. In order to use broadcasting network and communication network and synchronize the layers divided by SHVC, MMT standard is used [4]. MMT is the next generation transmission protocol, IP base transmission protocol. According to the MMT standard, video encapsulated and transmitted in a packet format, so that IP-based broadcasting and communication networks can be used simultaneously, and the receiver side can synchronize and play 8K UHD content. We propose an encapsulation method that can efficiently transmit and receive 8K UHD. The proposed method increases a payload rate and decreases an initial delay at the receiver. We experimentally verify that the efficiency is improved in terms of payload rate and initial delay at the receiver.

2. Broadcasting System for 8K UHD Services

2.1 Transmitting and Receiving Scheme using M-PLP and Scalable HEVC

In this section, an example broadcasting system for 8K UHD is described. The broadcasting system consists of 8K UHD contents provider, MMT generator, DVB-T2 (Digital Video Broadcasting-2nd Generation Terrestrial) multiplexer, modulator, HTTP streaming server and receivers supporting various resolutions from 2K to 8K. **Fig. 1** shows the service architecture for UHD broadcasting over heterogeneous networks.

For playing UHD, contents could be divided into 4K sub-screens because 8K UHD TV has not been commercialized yet. 4K UHD streams of each area are hierarchically divided into 3

layers, Layer0 (960x540) Layer1 (1920x1080) and Layer2 (3840x2160), using SHVC technology. Layer0, Layer1, signaling messages and audio are transmitted over DVB-T2 broadcasting system after multiplexing and Layer2 is transported over communication networks [5,6]. In this way, users can view various resolution content depending on the user environment.

Layer0, Layer1, audio and signaling messages are encapsulated on MMT packets and then transmitted to DVB-T2 multiplexer. DVB-T2 multiplexer uses M-PLP (Multiple - Physical Layer Pipeline) and then transmits the multiplexed data to Modulator. M-PLP is designed as 5 PLPs on the multiplexer. 5 PLPs are composed of 2 video contents and 3 data files. One of the videos consists of Layer0, audio and signaling message. The other consists of Layer1. The three data files consist of XML data including program information, JPG image of broadcast program, and XML data describing MMT Server URL information of Layer2.
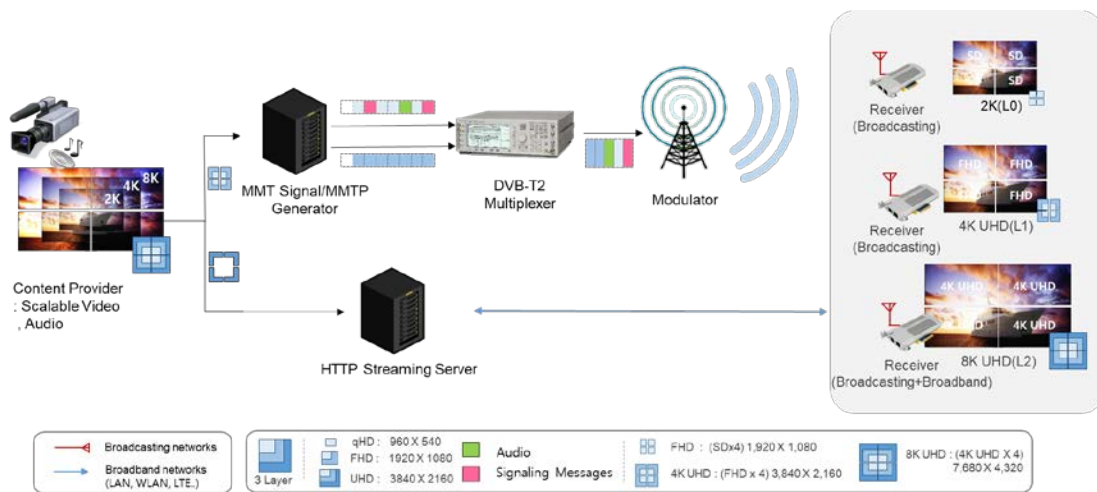

Fig. 1. Service architecture of 8K UHD broadcasting

2.2 Video stream format for transmitting based on MMT

The format in which the 8K UHD contents are transmitted through the broadcasting and the communication networks is defined by the MMT standard. MMT is a multimedia transmission protocol standard that can transmit contents based on IP. MMT defines an MPU (Media Processing Unit) based on ISO-BMFF, which is a logical structure that can independently consume the encoded media contents. MMT Generator encapsulates the encoded media content into the MPU structure for transmittng and consuming. MPU can be encapsulated in consideration of the smallest unit needed to be played. The information required for decoding the encoded media is stored with metadata such as 'ftyp', 'mmpu', 'moov', 'moof'. The encoded media stream is stored after the 'mdat' metadata. As shown in the **Fig. 2**, An 8K UHD content is divided into 4 areas named as Area # 1 - Area # 4. Each area is divided into three layers, and each layer is encapsulated in MPU. Totally, An 8K content is generated in 12 video streams, each of which is encapsulated in MPU.

Transmission scheduling method according to area informationis studied in [2]. It is proposed that MPUs are generated into MMTP packets and the packets included data of the same layers in each area are scheduled in the area-numerical order. **Fig. 3** shows that

transmission architecture for 8K. MPU Generator generates signaling messages and MPUs by the video stream like **Fig. 2**, and then signaling messages and MPUs are packetized MMTP packets. The signaling message includes required information for transmitting and consuming the MPUs. Also signaling message includes CI component and Group Information. Group Information offer information for distinguishing video streams corresponding to Area#1 to Area#4. CI component includes position information and time schedule of contents.
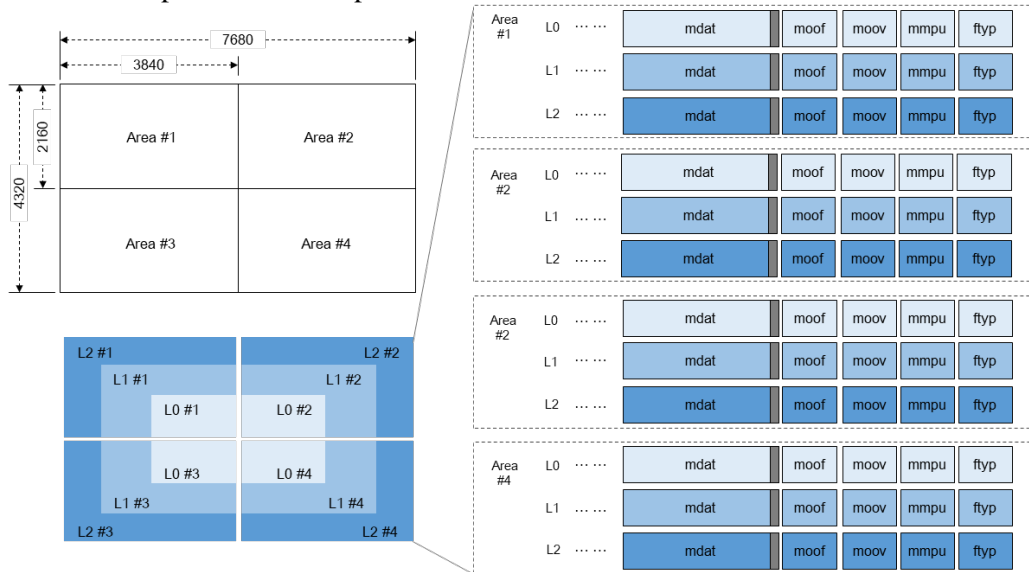

Fig. 2. 8K UHD divided into 4 areas and MPU strucure of video stream

When the MMTP packets are received at the UHD receiver, in order to play 8K UHD contents at the receiver, it is necessary to verify the information of the 12 MPUs received and compare the information about each layers, areas and time. The receiver have to check whether each area is transmitted without being broken, and then merge the 4 areas of the same layer into the 8K UHD video stream. The receiver repeat the checking and grouping the MPUs for passing to the SHVC decoder. The more the task to be compared per MPU, the greater the burden on the receiver. Therefore, this paper proposes a method to reduce this procedure.
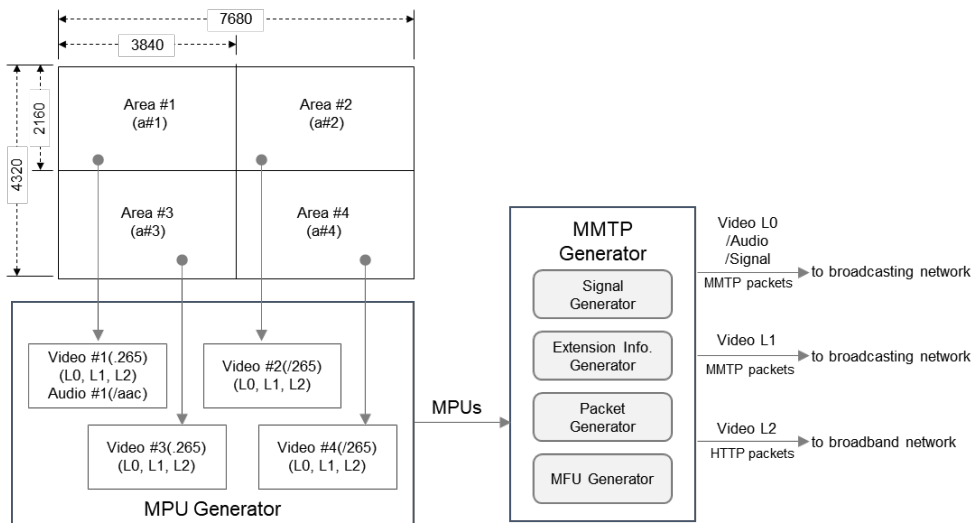

Fig. 3. MMTP generation and transmission system for 8K UHD

3. Proposed encapsulation method based on MMT

3.1 Proposed MPU structure

  We propose a method of encapsulating MPU including Area information. In the method described in Chapter 2, the 8K content is encapsulated to 12 MPUs. However, when MPU generator encapsulates 8K content in the proposed method, the MPUs arrange the frame of each area-numerical order as shown in **Fig. 4**. The left side of **Fig. 4** shows that the video streams arranged in frame order, while the right side presents that one MPU consist of the frames that arranged in area-numerical order. Because MMTP packets are generated by the proposed MPU structure, it is not necessary to check whether the received MPUs are the same area or not at the receiver. By including the frames of the same group areas in one MPU, the receive procedure is simplified. In addition, the overhead of MPU by its header can be reduced so that the payload rate is increased.
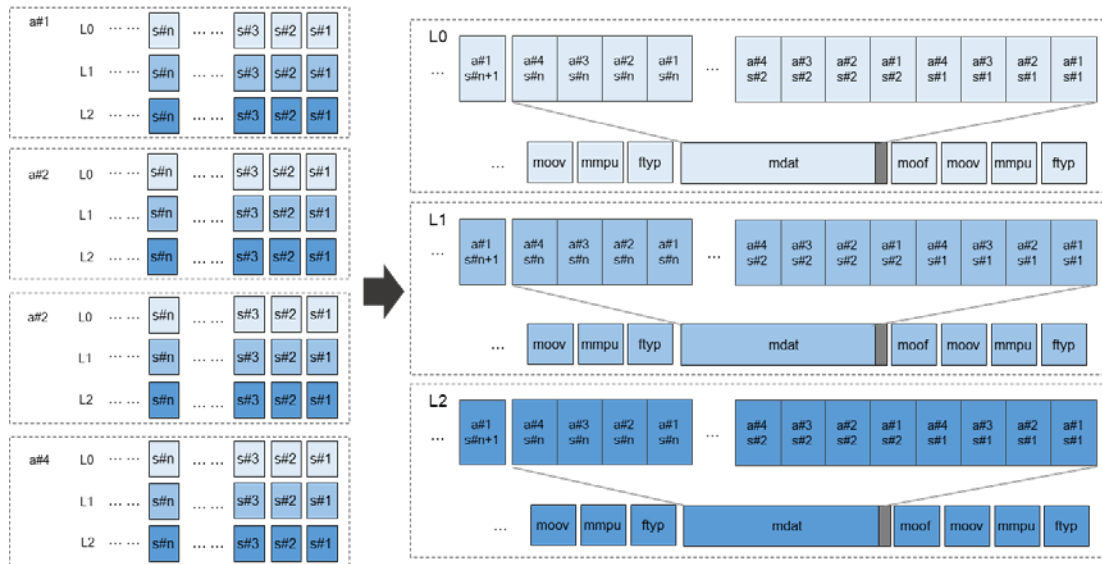

Fig. 4. Proposed MPU structure for 8K UHD

3.2 Information for distinguishing area in MPU structure

  For applying the proposed method, area information should be added into a hint track of MPU structure. **Fig. 5** shows the hint track including the area information. The hint track is the frame information in MPU. In the existing hint track, *'multiarea_flag'*, *'area_count'*, and *'area_length'* information added. *'Multiarea_flag'* is flag information that indicates whether the content divided area or not. *'Area_count'* is information indicating how many screens divided. *'Area_length'* is the each frame length in MPU. Through the *'area_length'* information, the frame of each area can distinguished. The reason for distinguishing them is that the frame of each area should be merged with the upper layer frames.

```
aligned(8) class MMTHSample {

   unsigned int(32) sequence_number;
  if (is_timed) {
   signed int(8)      trackrefindex;
       unsigned int(32) movie_fragment_sequence_number
     unsigned int(32) samplenumber;
     unsigned int(32) offset;
     unsigned int(8)  priority;
     unsigned int(8)  dependency_counter;
     unsigned int(32) offset;
     unsigned int(32) length;
     multiLayerInfo();
  } else {
     unsigned int(16) item_ID;
   }
}
```

```
aligned(8) class multiLayerInfo extends Box('muli') {
   bit(1)  multilayer_flag;
   bit(1)  multiarea_flag;
   bit(6)  reserved0;

  if (multilayer_flag==1) {
       bit(3)  dependency_id;
       bit(1)  depth_flag;
       bit(4)  reserved1;
       bit(3)  temporal_id;
       bit(1)  reserved2;
       bit(4)  quality_id;
       bit(6)  priority_id;
       bit(10) view_id;
   }
   else{
       bit(6) layer_id;
       bit(3) temporal_id;
       bit(7) reserved3;
   }
  if(multiarea_flag==1) {
       unsigned int(8)   area_count
       unsigned int(32*area_count)  sample_length;
   }
}
```

Fig. 5. An example of area information in MPU hint track

4. Comparison of efficiency

In this section, we compare the differences between the original method and the proposed method. There are two kinds of efficiency gap between the existing method and the proposed method. The first is the payload rate. The payload rate depends on the size of MPU transmitted in the MMTP packet. The second difference is the receiving procedure. Before we compare the difference of the receiving procedure, we describe the receiver structure and the requirements for playing the 8K UHD.

4.1 Efficiency of payload rate

The payload rate is defined as the total transmission size versus the MPU size like equation (1). If the encapsulated MPU is considered as a payload, the information, which is not the payload, is a periodic signaling messages and NULL packets. There are ways to reduce signaling messages and NULL packets to improve the efficiency of the payload. Signaling messages have to be sent periodically because they contain the information required to play the video stream at the receiver. Since the efficient period of signaling messages proposed in [2], we do not compare the efficiency of scheduling of signaling messages.

$$P_r = \frac{total\ MPU\ size}{total\ MMTP\ Packet\ size} \qquad (1)$$

The video stream is encapsulated into the MPU and then packetized into the MMTP packet. At this time, the MPU size does not match the MMTP packet size. When the MPU transmitted as the MMTP packets and the rest of the MPU is smaller than the MMTP packet size, the MMTP packet filled with NULL data. According to the MMTP transmission rule, it can see that the smaller the MPU size, the lower the payload rate. However, the efficient size of the MPU does not fixed because the payload rate may depend on the characteristics of the video. However, existing encapsulation method for 8K UHD transmission generates up to 12 MPUs per a frame according to the area and layer. The proposed encapsulation method generates up to 3 MPUs per the frame according to the layer. Reducing the number of MPUs per content

and the number of null packets, thus we improve the payload rate. The payload rate is of the proposed method verified through the experiment of Chapter 5.

4.2 Receiving requirements

The receiver structure for 8K UHD is based on the structure of 8K UHD broadcasting systems in section 2[7]. The receiver structure shown in the **Fig. 6** and the operation process is as follows. MMT-based receiver receives MMTP packets and MPUs through IP receiver and HTTP receiver. The IP receiver extracts MMTP packets from IP packets, which are de-multiplexed from the M-PLP on the broadcasting networks. The MMTP packets extracted through the MMTP Parser and the MFU parsers.

The 8K UHD decoder arranges sub-frames as shown in the **Fig. 7**. The decoder works on each area and each thread simultaneously. All decoders are synchronized based on the timestamp. SHVC decoders are decoded based on the lower Layer. Therefore, if there is no lower layer, the upper layer is not decoded. In addition, if the video stream of one area is broken, each sub-frame is not decoded.
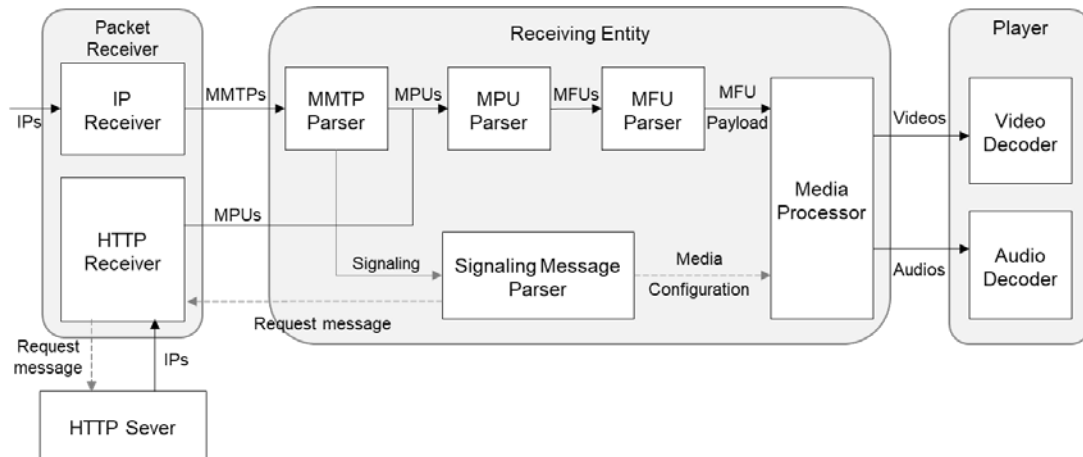

Fig. 6. Receiver structure of hierarchically encoded video based on MMT

The requirements for the receiver structure to play8K UHD video stream to play according to the receiver and the 8K UHD decoder structure are as follows:

- The information of the received stream should include frame order information and timestamp.
- Area information of the video stream should be included.
- Layer information and the video stream length of each layer should be included.
- Any stream should not be broken.
- The upper layer cannot be decoded, if the lower layer video stream is broken.
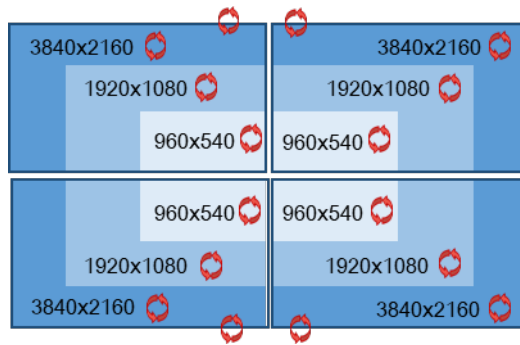
Fig. 7. The structure of scalable HEVC decoder for 8K UHD

4.3 Compare of receiving procedure

The receiving procedure of conventional method is shown in **Fig. 8** (a). The process of receiving 8K MMTP and delivering it to the decoder is as follows. When MPU and MFU header are parsed, the timestamp and Group_Info of a sub-frame have to equal to those of the other sub-frames.
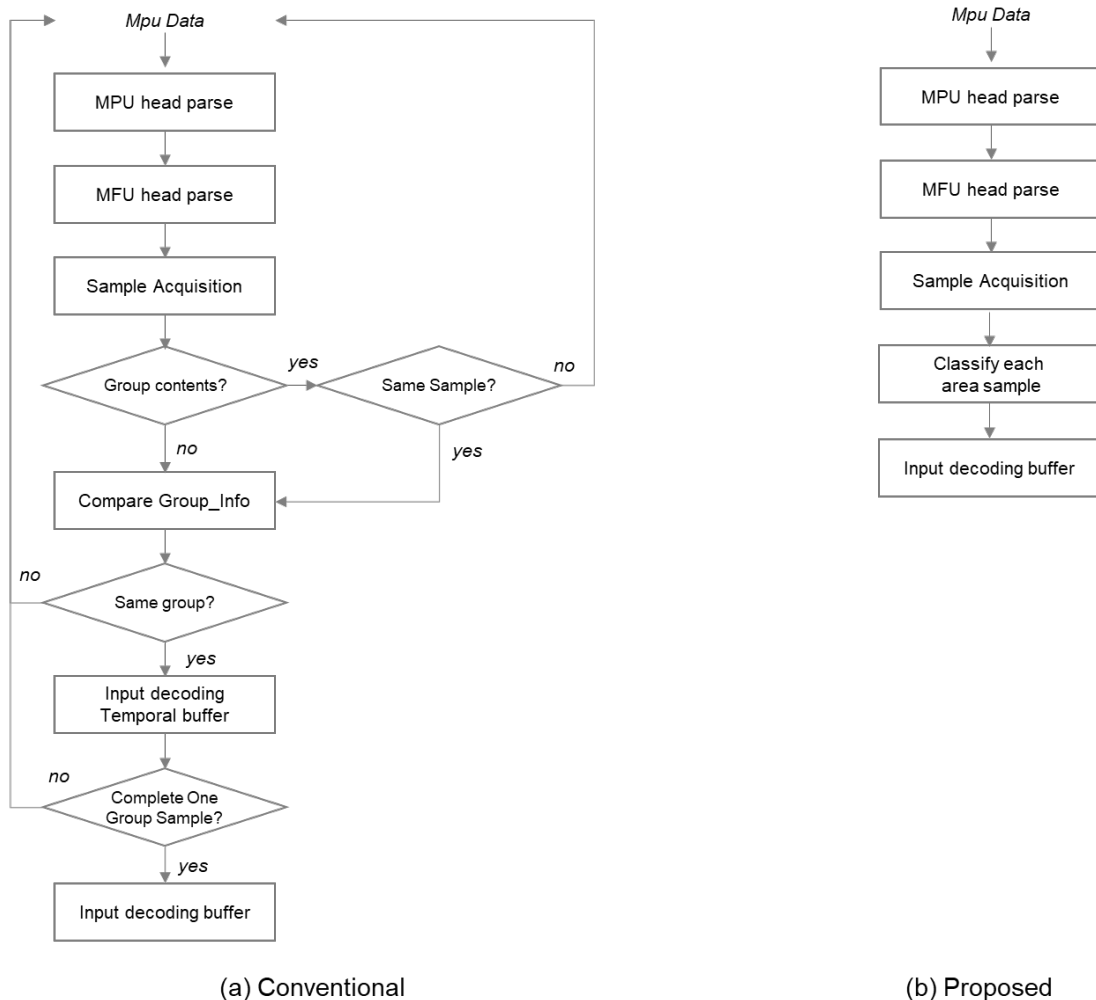

(a) Conventional   (b) Proposed

Fig. 8. Comparison of receiving procedure

The receiving procedure of proposed method is shown in **Fig. 8** (b). The process of verifying the frame information among the frames belonging to the same group is simplified. Because the frames belonging to the same group are included in MPU, so the proposed scheme can omit the process checking with the timestamp. Unless some MMTP packets are lost, the video streams of the same group are received in the MPU and then passed to the decoder.

5. Experiment

In this section, the performances of the proposed method are evaluated and compared with the conventional method. **Figure 9** shows the MMT-based 8K transceiver system used to verify the proposed encapsulation efficiency. **Figure 9** (a) is the transmission system, which consisted of the MMT-based packet generator, the broadcasting signal generator, and HTTP server serving enhanced layer contents. **Figure 9** (b) is the reception system, which is consisted of 4 monitors with 4K resolution.

(a) Transmission system

(b) Reception system

Fig. 9. 8K UHD Broadcasting system

The 8K UHD contents used in the experiment are as called "Ullengdo" and "satellite", respectively. The contents split into 4 areas before encoding, and each divided video stream is encoded according to the configuration information described in the **Table 2**.

6. Results

  This section shows the test results of the proposed method to measure how effective it is on the receiving when compared with the conventional method. **Figure 10** shows that 8K UHD contents are displayed through 4 monitors with 4K resolution. 8K contents transmit after being divided into 4 4K contents, and then each part of 4K monitors play it on 4K resolution after decoding.
  We measure the initial delays of the proposed method and the conventional method and summarize them in **Table 3.** From **Table 3**, one can see that the initial delay of the proposed method is a half of the conventional method. **Table 4** and **Fig. 11** show the payload rate according to the size of MPU and MMTP. We only experiment the Layer0 and Layer1, because Layer2 is uploaded on HTTP streaming server, so that Layer2 does not be packetized to MMTP packet. Depending on characteristics of video stream, the payload rate is verifying however, the performance of the payload rate is improved in the proposed method.

Table 2. Input configuration for experimental

Video Title	Ullengdo		Satellite
Media Play Time	116s		78s
MPU duration	1s		
QP Level	32		
The Number of Area	4		
Audio(AAC)	Sampling Rate	44.1khz, 2Ch : Broadcast(PLP0)	
Video(SHVC)	Frame Rate	30fps	
	Resolution	960x540(L0) : Broadcast(PLP0)	
		1920x1080(L1) : Broadcast(PLP1)	
		3840x2160(L2) : HTTP	
	Intra period	32	

Table 3. The comparison result of experiment

	Ullengdo		Satellite	
	Conventional	**Proposed**	**Conventional**	**Proposed**
Initial delay time(sec)	4.5	2.3	4.6	2.2

Fig. 10. The result of experiment

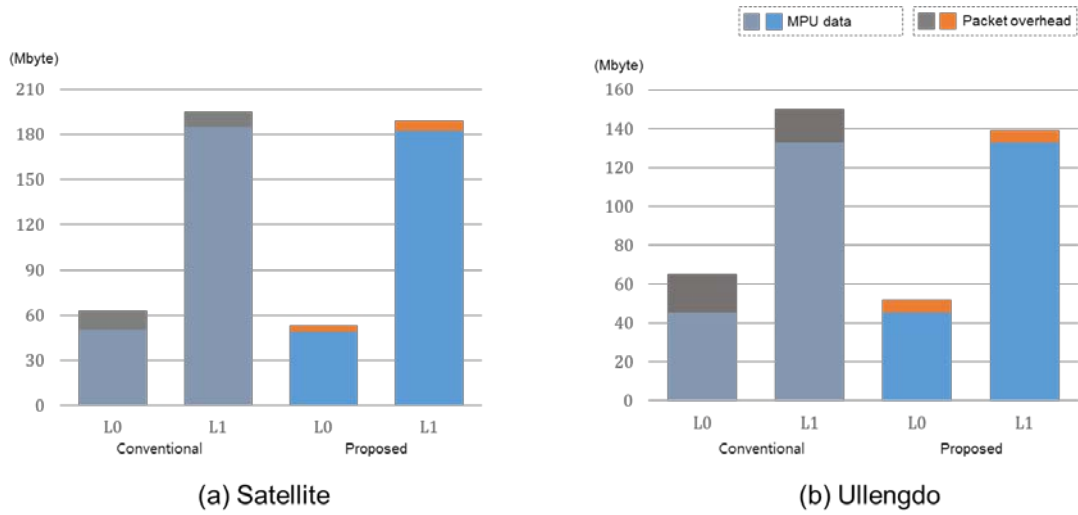(a) Satellite   (b) Ullengdo

Fig. 11. The payload rate according UHD contents

Table 4. The payload rate according UHD contents

			MPU(byte)	MMTP(byte)	Payload rate (%)	
Ullengdo	Conventional	L0	50,435	62,938	80.1	87.7
		L1	185,545	194,408	95.4	
	Proposed	L0	49,463	52,977	93.3	94.9
		L1	182,473	189,025	96.5	
Satellite	Conventional	L0	45,033	65,835	68.4	78.5
		L1	132,995	150,106	88.6	
	Proposed	L0	45,033	49,880	90.2	92.8
		L1	132,995	139,435	95.4	

7. Conclusions and Further Works

Some research of the 8K UHD broadcasting systems were proposed. However, transmitting and receiving method for 8K UHD on previous works are not efficient. In this Paper, we proposed the 8K UHD MMT-based encapsulation method for efficient transmission and reception. 8K UHD stream rate is beyond the range that the current broadcasting network can provide. In order to transmit 8K UHD within the limited transmission bitrate of broadcasting network, 8K UHD contents are encoded by SHVC and then transmitted over heterogeneous networks. In addition, the receiver is implemented with four 4K UHD display instead of 8K UHD display. In order to use broadcasting and communication networks simultaneously, MMT standard is applied on the proposed method. For increasing efficiency, we redesign the MPU structure that reduces the initial delay time and increases the payload rate. We experimentally verify that the efficiency is improved in terms of payload rate and initial delay at the receiver.

References

[1] Y. Miki, T. Sakiyama, K. Ichikawa, M. Abe, S. Mitsuhashi, M. Miyazaki, A. Hanada, K. Takizawa, I. Masuhara and K. Mitani, "Ready for 8K UHDTV Broadcasting in Japan," *IBC2015 Conference*, September 10-14, 2015. Article (CrossRef Link).

[2] Y. Sohn, M. Cho, and J. Paik, "Design of 8K Broadcasting System based on MMT over Heterogeneous Networks," *KSII Transactions on Internet and Information Systems*, Vol. 11, No.8, pp.4077-4091, August, 2017. Article (CrossRef Link).

[3] ISO/IEC 23008-2, "Information technology – High efficiency coding and media delivery in heterogeneous environments – Part2 : High efficiency video coding," 2015. Article (CrossRef Link).

[4] ISO/IEC 23008-1, "Information technology – High efficiency coding and media delivery in heterogeneous environments – Part1 : MPEG media transport(MMT)," 2014. Article (CrossRef Link).

[5] Y. Ryu, K. Park, J. Wee, and K. Kwon, "An Efficient 4K and 8K UHD Transmission Scheme on Convergence Networks with Broadcasting and LTE by using Coordinated Multi-Point Transmission System," *KSII Transactions on Internet and Information Systems*, vol. 11, no.8, pp. 4092-4104, August, 2017. Article (CrossRef Link).

[6] J. Oh, Y. Won, J. Lee, Y. Kim, J. Paik and J. Kim, "A Study of Development of Transmission Systems for Next-generation Terrestrial 4K UHD & HD Convergence Broadcasting," *Journal of Broadcast Engineering*, Vol.19, No.6, pp.767-788, November 2014. Article (CrossRef Link).

[7] S. Song, J. Bang and J. Paik, "Design and Implementation of Receiver's Buffer Model for Stably Receiving a Hierarchically Encoded UHD content Based on MMT," *Korean Institute of Broadcast and Media Engineers*, Vol. 22, No. 5, September 2017. Article (CrossRef Link).

Seulki Song received the B.S degree in Multimedia from Seoul Women's University, Seoul, Korea, in 2015 and M.S. degree in Information Media from Seoul Women's University in 2017. She has been an assistant researcher in Korea Electronics Technology Institute (KETI), Seoul, Korea, since 2017. Her research interests are digital broadcasting and digital communications system.

Youngsu Ryu received the B.S. degree in Information & Communication Engineering, Sejong University, Seoul, Korea, in 2015. He has been an assistant researcher in Korea Electronics Technology Institute (KETI), Korea, since 2014. His research interests are digital broadcasting, energy storage system, and cooperative communication.

Jungwook Wee received the B.S. and M.S. degrees in Electrical Engineering from Chung-Ang University, Seoul, Korea, in 1999 and 2001, respectively. He is currently working toward the Ph.D. degree in the school of electrical and Electronics Engineering from Chung-Ang University. Since 2001, he has been with Contents Convergence Research Center in Korea Electronics Technology Institute (KETI), Seoul, Korea. His research interests are in the areas of signal processing for digital communication and wireless/wired communication systems.
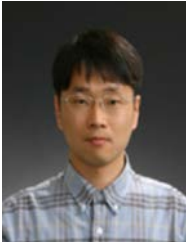
Kyungwon Park received the B.S. and M.S. degrees in electrical engineering from Chung-Ang University, Seoul, Korea, in 1999 and 2001, respectively, and Ph.D. degrees in the school of electrical and electronic Engineering from Chung-Ang University in 2005. Since 2005, he has been working at Korea Electronics Technology Institute (KETI), Seongnam-si, Gyeonggi-do, Korea, where he is presently the managerial researcher of Smart Network Research Center. His research interests are in the areas of wireless/wired communications system design and multiple antenna systems including multiple-input multiple-output systems and smart antenna systems.

Kiwon Kwon received B.S. and M.S. degrees in computer engineering from Kwangwoon University, Korea, in 1997 and 1999, He also received the Ph.D. degree in the School of Electrical & Electronics Engineering from Chung-Ang University, Korea, in 2011. In 1999, he joined in KETI, Korea, where he is presently a principal researcher at the Smart Network Research Center. His research interests are in the area of and advanced broadcasting/communication system and energy convergence networks..