ATSC 3.0 ROUTE/DASH Signaling for Immersive Media: New Perspectives and Examples

DONGHO YOU1, SUNG-HOON KIM2, AND DONG HO KIM3, (Senior Member, IEEE)

1Department of Information and Communication Engineering, Hannam University, Daejeon 34430, South Korea
2Electronics and Telecommunications Research Institute (ETRI), Daejeon 34129, South Korea
3Department of Electronic and IT Media Engineering, Seoul National University of Science and Technology, Seoul 139743, South Korea

Corresponding author: Dong Ho Kim (dongho.kim@seoultech.ac.kr)

This work was supported in part by the Korean Government through the Electronics and Telecommunications Research Institute (ETRI) Grant (21ZH1200, the research of the basic media contents technologies), and in part by the Korean Government [Ministry of Science and ICT (MSIT)] through the National Research Foundation of Korea (NRF) under Grant 2018R1D1A1B07050509 and Grant 2021R1G1A1091541.

ABSTRACT This paper examines the protocol structure of ATSC 3.0, the next-generation terrestrial broadcasting standard, and explains in-depth the technology of Real-Time Object Delivery over Unidirectional Transport (ROUTE) and MPEG Dynamic Adaptive Streaming over HTTP (DASH). Specifically, it considers both low and service signaling methodologies of ROUTE/DASH, that enable immersive media, stereoscopic 3D, and 360 virtual reality (VR), delivery in ATSC 3.0. Furthermore, this paper also provides the signaling examples to give readers new perspectives on immersive broadcasting through ROUTE/DASH in ATSC 3.0. In the era of COVID-19 pandemics, this kind of type, the ATSC 3.0 immersive media service method is expected to become an alternative to performance and media consumption in the future.

INDEX TERMS ATSC 3.0, ROUTE, MPEG-DASH, signaling, immersive media, 3D video, 360-degree video, virtual reality.

I. INTRODUCTION

ATSC 3.0 is the first fully IP-based broadcasting standard that revolutionizes TV broadcast delivery [2], [3]. It is designed for flexible receptions on both fixed and mobile devices with high quality and interactivity. To this end, ATSC 3.0 uses Real-Time Object Delivery over Unidirectional Transport (ROUTE) [4] and MPEG Dynamic Adaptive Streaming overHTTP (DASH) [5] for both broadcast and broadband delivery of media data. ROUTE is an extension of File Delivery over Unidirectional Transport (FLUTE) [6] in order to embrace MPEG-DASH, which splits the media data into multiple segments according to ISO Base Media File Format (ISOBMMF) [7]. It allows discovering and accessing available services and their segmented contents by specific signaling, which describes essential characteristics that make up each service such as content type and video coding format. Details are described in ATSC A/331 Standard [8].

The basic functions of the ATSC 3.0 standard have been currently completed but the impact of future broadcast technology and countermeasures is still being discussed. The discussion can be largely divided into (i) future video compression technology and (ii) future video formats and services. The former is about identifying the development status and advantages of new video compression techniques such as Versatile Video Coding (VVC), Essential Video Coding (EVC), and Low Complexity Enhancement Video Coding (LVEVC). The latter is about identifying new video formats such as 8K, High Frame Rate (HFR), point cloud, immersive media, i.e., Free-viewpoint TV (FTV), Augmented Reality (AR), and Virtual Reality (VR).

In the case of South Korea, major events of the 2018 Pyeongchang Winter Olympics were broadcast live on terrestrial Ultra High Definition (UHD) broadcasting for the first time in the world. In the case of the United States, the Federal Communications Commission (FFC) approved the use of ATSC 3.0 in November 2017, approximately 62 Designated Market Areas (DMAs) are currently building infrastructures that allow more than 75% of viewers to receive ATSC 3.0 broadcasts. In addition, as the demand for high-quality media increases, at Consumer Electronics Show (CES) 2020, various 8K-UHD ATSC 3.0 receivers were exhibited and video streaming-related technologies for 8K-UHD services were introduced. In addition, many researchers are recently...
developing core technologies to provide seamless convergence services by interconnecting ATSC 3.0 broadcasting and 5G communication networks. It is expected that this effort can be utilized in the future for high-quality, immersive media, mobile, and additional services.

In this paper, to fit into the new era, we consider immersive broadcasting in ATSC 3.0. Specifically, we deal with ATSC 3.0 ROUTE/DASH and its comprehensive perspective on the signaling methods, i.e. Low-Level Signaling (LLS) and Service Layer Signaling (SLS), for immersive broadcasting. The former is for the discovery of all services provided to users and the latter is for the access and presentation of a specific service. We also provide new perspectives on upcoming broadcast services by concentrating on immersive broadcasting and its specific signaling methods. Finally, since ROUTE/DASH uses a media presentation description (MPD) for the service presentation which varies according to the content formats, we also present some examples of MPD signaling that enables the immersive service via ROUTE/DASH in ATSC 3.0.

The remainder of this paper is organized as follows. Basic principle including LLS and SLS of ATSC 3.0 ROUTE protocol is described in Section II. Various perspectives and examples about immersive broadcasting and MPD signaling are discussed in Section III and IV, respectively. Finally, the conclusion is given in Section V.

II. BASIC PRINCIPLE OF ROUTE/DASH IN ATSC 3.0

Fig. 1 shows a conceptual ATSC 3.0 ROUTE protocol stack in which MPEG-DASH segments can be delivered through both ROUTE/UDP/IP and HTTP/TCP/IP. Native broadcasting services are generally delivered through ROUTE/UDP/IP and additional multimedia contents, such as secondary language audio and text, are delivered through HTTP/TCP/IP along with the native services. In addition, non-real-time (NRT) contents, such as the electronic service guide (ESG) and emergency alert (EA), can be also delivered through both ROUTE/UDP/IP and HTTP/TCP/IP. The ROUTE-specific SLS is delivered through ROUTE/UDP/IP, while bootstrap signaling information is provided by means of the service list table (SLT). The SLT is one of LLS and supports building a basic service list, as well as discovering the SLS and accessing each ATSC 3.0 service. The details are in [8]. The relationship between SLT and SLS is illustrated in Fig. 2, and the brief process is described as follows.

1) ATSC 3.0 users first receive SLT which provides the bootstrap information to discover the SLS for each service. The SLT includes the IP addresses/ports of each layered coding transport (LCT) channel that carries an SLS. In general, a ROUTE session may consist of one or more LCT channels and each LCT channel carries a content component, such as video, audio, and caption.

2) Before acquiring the delivered contents by broadcasters, ATSC 3.0 users have to parse the SLS that is comprised of XML-based fragments: the user service bundle description (USBD), the service-based transport session instance description (S-TSID), and the DASH media presentation description (MPD). These fragments organizationally provide information about the characteristics of the desired service.

3) The USBD fragment of the ATSC 3.0 is similar to that of the 3GPP MBMS [12]. It provides overall information on the service such as service identification (ID) and device capabilities (e.g., HD or UHD). Especially, the USBD refers to MPD and S-TSID fragments to access a chosen service and content components, respectively.

4) The MPD fragment provides identification information for the content components that make up the service, such as the content ID and type (e.g., video or audio), video coding format, etc. [13]. Therefore, user devices determine how to decode and play the contents according to the MPD.

5) The S-TSID fragment provides the information on where to acquire the DASH segments described in the MPD. In other words, the S-TSID tells users which LCT session delivers the DASH-formatted content components. Importantly, the ContentInfo of the S-TSID and the Representation ID of the MPD must be matched.

III. PERSPECTIVES ON IMMERSIVE BROADCASTING

In this section, we deal with hybrid immersive broadcasting services, in which various qualities of the immersive media can be provided via only one physical transmission simultaneously. Firstly, we address a hybrid stereoscopic 3D service, in which low-quality 2D video, and high-quality 2D and 3D videos can be provided simultaneously for both mobile and fixed TV users. Secondly, we address a tile-based hybrid stereoscopic 360 VR service, in which low-quality mono video, high-quality mono video, and stereoscopic 360 VR video can be provided simultaneously for both mobile and fixed VR users. This can be considered as an extension of the first one that enables an adaptive delivery of 360 VR videos.

A. FIXED/MOBILE HYBRID 3D SERVICE

Fig. 3 shows a concept of a hybrid 3D service for fixed and mobile users in which the UHD left-view and the HD
right-view videos are provided. Since mobile users generally experience worse channels than fixed users, low-quality HD right-view (mono) video is only decoded. Conversely, fixed users decode both HD right and UHD left-view videos because of better channel condition, so high-quality UHD right-view (mono) and stereoscopic 3D services can be provided in accordance with the device’s capabilities. In addition, the fixed/mobile hybrid UHD 3D service can be divided into two cases according to the video coding formats that are used as follows: the HEVC-based simulcast and the SHVC-based layered services. The H.265/HEVC is a video compression standard that is an extension of the H.264/MPEG-4 AVC [9]. The HEVC supports resolutions up to 8192 × 4320, including 8K UHD. The SHVC supports scalability in spatial, SNR, and color terms, and thus it can reduce the average data rate by up to approximately 15% for a 4K UHD video [10]. The details by video coding methods are as follows:

1) **HEVC-BASED SIMULCAST SERVICE**

The HEVC-based simulcast 3D system is illustrated in Fig. 4, in which the UHD left-view and the HD right-view of a stereoscopic 3D video are encoded independently by two HEVC encoders. The HD right-view for mobile users is modulated by the low-order modulation, such as the QPSK and 16QAM for high reliability, whereas the UHD left-view is modulated by the high-order modulation 256QAM due to its large amount of data. The two modulated video streams are delivered by different physical layer pipes (PLPs) and provide UHD and HD 2D services for the fixed and mobile devices, respectively. The number of PLPs used can vary...
2) SHVC-BASED LAYERED SERVICE

Fig. 5 shows that the HD right-view and the UHD-left view videos are encoded into a base layer (BL) and an enhancement layer (EL), respectively. The UHD right-view is first scaled down into HD quality and encoded into a BL. Next, the down-scaled HD right-view is again scaled up into the original UHD quality, compared with the UHD left-view to obtain the inter-layer prediction value, and re-encoded into an EL. The EL is independently undecodable without the BL because it is encoded on top of the BL. Hence, the BL needs to be modulated by the low-order modulation for the purpose of its high reliability, while the UHD EL is modulated by the high-order modulation due to its large amount of data. The BL and EL are also delivered through different PLPs, respectively. For example, PLP #1 is for the BL and PLP #2 is for the EL. For the mobile HD service, the BL is sufficient, whereas the fixed devices require both the BL and EL to reconstruct the UHD 2D content. Furthermore, if user devices have the capabilities to display stereoscopic 3D video, the devices can combine the two video views to display stereoscopic 3D. This is unlike the HEVC-based simulcast 3D service in which one video encoder/decoder, such as SHVC encoder/decoder [11], is only required to provide users with three services simultaneously, including HD 2D service, and UHD 2D/3D services.

B. TILE-BASED HYBRID 360 VR SERVICE

As the demand for immersive media grows, various types of VR services such as 360 VR media are emerging. According to [14], providing the stereoscopic 360 VR is very important for making the VR experience more immersive. To this end, the left 360 and right 360 videos must be separately projected and displayed to each eye. However, because poorly implemented stereoscopic 360 VR service can cause major discomforts, such as headaches, eye strain, and nausea, many studies have been widely conducted to reduce this problem.

Fig. 6 shows an overview of the tile-based hybrid stereoscopic 360 VR service, in which both the left 360 and right 360 VR videos are equirectangularly projected and separated into independent 16 tiles (4 × 4). The number of tiles can vary sufficiently depending on the purpose. These tiles are used for parallel processing. Similar to the SHVC-based layered service, the 8K (7680 × 4320) 360 right video is first scaled down into low resolution 360 video (e.g., QHD, 2560 × 1440), and encoded by a single SHVC encoder along with the original 8K 360 left video. This encoding process is identical to the conventional SHVC encoder shown in Fig. 5. The encoded low-quality (360 right) and high-quality (360 left) streams are also called BL and EL, respectively. The BL is for VR users within a bad channel environment (e.g., mobile VR), whereas the EL is for VR users within a good channel environment (e.g., fixed VR). In addition, if VR users are capable of displaying stereoscopic 3D within a good network environment, both the BL and EL are combined to provide stereoscopic 360 VR videos for the user.

IV. MDP SIGNALING EXAMPLES

As we described in Section II, ROUTE/DASH-specific signaling includes three fragments: USBD/USD, S-TSID, and MPD. However, for the immersive broadcasting and its signaling considered in this paper, it is sufficient to redefine the MPD fragment only. This is because USBD/USD and S-TSID are not interested in content-related information such as codec, frame rate, resolution, view position, dependency, etc. In this section, therefore, we present some examples of MPD signaling for hybrid immersive broadcasting.

A. HEVC-BASED SIMULCAST SERVICE

As mentioned in Section III-A1, the HD right-view and the UHD left-view videos are independently encoded. The view position information must be described in the MPD fragment so that the ATSC 3.0 users identify which view is left or right. The MPD signaling for the HEVC-based simulcast service is described in Fig. 7, in which some elements and attributes are omitted for simplicity. There are two AdaptationSet elements that separately have a different
B. SHVC-BASED LAYERED SERVICE

In the case of the SHVC-based layered service, the HD right-view and the UHD left-view are encoded into the BL and EL, respectively. Because the EL is encoded on top of the BL, EL is independently decodable without the BL. In other words, there is a dependency between the BL and EL, so the MPD signaling must indicate the dependency along with the stereo view position. The example of the MPD signaling for the SHVC-based layered service is described in Fig. 7(b). Like the case of the HEVC-based 3D service, there are two video components (i.e., AdaptationSets) with their own qualities (i.e., Representations). The first AdaptationSet contains the BL and indicates the HD right-view video stream. It is for mobile devices that are experiencing bad channel conditions. On the other hand, the second AdaptationSet contains the EL, and it is decoded into the UHD left-view video with reference to the BL that is contained in the first AdaptationSet. To indicate the dependency (or layering), the Representation ID “rv1” of the first AdaptationSet must be indicated in the @dependencyId attribute in the Representation of the second AdaptationSet.
decoding of the UHD left-view video, both AdaptationSets must be chosen. The combination of the first and second AdaptationSets is for the 3D service, and the stereoscopic mapping is indicated by EssentialProperty, such as in the case of the HEVC-based 3D service.

### C. TILE-BASED HYBRID 360 VR SERVICE

Fig. 8 shows the example of MPD signaling that enables the tile-based hybrid stereoscopic 360 VR service. In general, spatial relationship description (SRD) is used to indicate the tile position in MPD fragment. According to [15], the position shall be signaled by using a @schemeIdUri with urn:mpeg:dash:srd:2014 and a @value of SupplementalProperty at an AdaptationSet. Especially, the @value must include following parameters in order: source_id, object_x, object_y, object_width, object_height, total_width, total_height and spatial_set_id. These denote an identifier for the original video, the horizontal and vertical positions of the top-left corner of the associated tile, the width and height of the associated tile, the width and height of the original video, and an identifier for a group of tiles, respectively. All parameters are non-negative integer in the decimal representation, and the last spatial_set_id is optional.

In Fig. 8, MPD of the tile-based hybrid stereoscopic 360 VR service is described. We consider an original 8K left 360° video that is separated into 16 tiles (i.e., the resolution of each tile is 1920 × 1080) and the @value of the first tile of left video is described as (2, 0, 0, 1920, 1080, 7680, 4320), in which assuming that the source_id of left video is “2”. Similarly, the @value of the second tile of left video can

![Diagram of MPD signaling for tile-based hybrid stereoscopic 360 VR service](image_url)
also be described as (2, 1920, 0, 1920, 1080, 7680, 4320). In addition to the spatial relationship of tiles, we should describe the stereoscopic relationship between the left and right videos, as well as the dependency between the BL and EL (refer to Section IV-A and IV-B).

V. CONCLUSION
This paper dealt with ATSC 3.0 ROUTE/DASH protocol and its specific signaling for immersive broadcasting. First, we presented service structures categorized in accordance with the video coding formats HEVC and SHVC. Especially in the case of using SHVC, three services, i.e., low-quality 2D, high-quality 2D/3D services, can be provided simultaneously through one physical channel. In addition, we presented MPD signaling examples that enable the aforementioned immersive broadcasting. We believe that the presented structures and signaling examples may give readers new perspectives on immersive broadcasting via ATSC 3.0 ROUTE/DASH protocol.

ACKNOWLEDGMENT
An earlier version of this paper was presented in part at the International Conference on Ubiquitous and Future Networks (ICUFN) [1] [DOI: 10.1109/ICUFN.2018.8437012].

REFERENCES
[1] D. You, E. Jeong, and D. H. Kim, “MPEG-DASH MPD for tile-based hybrid stereoscopic 360-degree video streaming,” in Proc. 10th Int. Conf. Ubiquitous Future Netw. (ICUFN), Prague, Czech Republic, Jul. 2018, pp. 682–684.
[2] Advanced Television Systems Committee (ATSC). ATSC 3.0 Standards. [Online]. Available: https://www.atsc.org/standards/atsc-3-0-standard/
[3] R. Chernock and J. C. Whittaker, “Next-generation broadcast television: ATSC 3.0 [standards in a nutshell],” IEEE Signal Process. Mag., vol. 33, no. 1, pp. 158–162, Jan. 2016.
[4] G. K. Walker, T. S. Stockhammer, G. Y.-K. M. Wang, and C. Lo, “ROUTE/DASH IP streaming-based system for delivery of broadcast, broadband, and hybrid services,” IEEE Trans. Broadcast., vol. 62, no. 1, pp. 328–337, Mar. 2016.
[5] I. Sodagar, “The MPEG-DASH standard for multimedia streaming over the internet,” IEEE Multimedia, vol. 18, no. 4, pp. 62–67, Apr. 2011.
[6] T. Paila, M. Luby, R. Lehtonen, V. Roca, and R. Walsh, FLUTE—File Delivery Over Unidirectional Transport, document IETF RFC 3926, Oct. 2004.
[7] Information Technology—Coding of Audiovisual Objects—Part 12: ISO Base Media File Format, Standard ISO/IEC 14496-12:2008, 2008.
[8] ATSC Standard: Signaling, Delivery, Synchronization, and Error Protection, document A/331:2017, Dec. 2017.
[9] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, “Overview of the high-efficiency video coding (HEVC) standard,” IEEE Trans. Circuits Syst. Video Technol., vol. 22, no. 12, pp. 1649–1668, Dec. 2012.
[10] Y. Ye and P. Andriov, “The scalable extensions of HEVC for ultra-high-definition video delivery,” IEEE Multimedia Mag., vol. 21, no. 3, pp. 58–64, Jul. 2014.
[11] H.-J. Yim, B.-M. Lim, B. Lee, N. Hur, and S. I. Park, “8K-UHD service platform using SHVC for ATSC 3.0-based terrestrial broadcasting,” in Proc. IEEE Int. Symp. Broadband Multimedia Syst. Broadcast. (BMSB), Chengdu, China, Aug. 2021, pp. 1–3.
[12] 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Multimedia Broadcast/Multicast Service (MBMS); Protocols and Codes, document 3GPP TS 26.346, Version 12.4.0, Dec. 2014.
[13] Guidelines for Implementation: DASH-IF Interoperability Point for ATSC 3.0 (Version 0.9), DASH Ind. Forum, DASH-IF Community Rev., Beaverton, OR, USA, Aug. 2016.
[14] M. Rowell. (Sep. 2015). Stereo vs Mono 360 Video for VR. [Online]. Available: http://360labs.net/blog/stereo-vs-mono-360-video-vr
[15] O. A. Niamut, E. Thomas, L. D’Acunto, C. Concolato, F. Denoaul, and S. Y. Lim, “MPEG DASH SRD: Spatial relationship description,” in Proc. 7th Int. Conf. Multimedia Syst. (MMsys), Klagenfurt, Austria, May 2016, pp. 1–8.

SUNG-HOON KIM received the B.S., M.E., and Ph.D. degrees in electronics engineering from Korea Advanced Institute of Science and Technology (KAIST), South Korea, in 1994, 1996, and 2008, respectively. Since 2000, he has been a Principal Researcher at the Electronics and Telecommunications Research Institute (ETRI), South Korea. His research interests include digital signal processing in the fields of terrestrial and mobile DTV, 3DTV, and UHDTV broadcasting systems.

DONG HO KIM (Senior Member, IEEE) received the B.S. degree from Yonsei University, South Korea, in 1997, and the M.S. and Ph.D. degrees from the Korea Advanced Institute of Science and Technology (KAIST), South Korea, in 1999 and 2004, respectively, all in electrical engineering. He was with the 4G Wireless Technology Laboratory, Samsung Advanced Institute of Technology (SAIT), from 2004 to 2006, and the Mobile Communications Research Institute, Samsung Electronics, from 2006 to 2007. Since 2007, he has been with the Department of Electronics and IT Media Engineering, Seoul National University of Science and Technology (SeoulTech), South Korea. His current research interests include the convergence of broadcasting and mobile communication: joint source-channel coding, multimedia signal processing, and all aspects of OFDM-MIMO systems.

DONGHO YOU received the B.S. and M.S. degrees in media IT engineering and the Ph.D. degree in broadcasting and communications engineering from the Seoul National University of Science and Technology (SeoulTech), South Korea, in 2012, 2014, and 2018, respectively. He was a Senior Researcher with the Deutsche Telekom Chair of Communication Networks, Technical University Dresden (TU Dresden), from 2018 to 2021. He is currently an Assistant Professor with Hannam University, South Korea. His current research interests include 5G communication networks and services for robust and reliable multimedia communications.