A Viewport Adaptive Low-latency Streaming System for Large-scale Multi-view Service

Jewon Lee, Eunhee Hyun, Soon-Choul Kim, and Joon-Young Jung
Media Broadcasting Research Section
Electronics and Telecommunications Research Institute
Daejeon, Repulbic of Korea
jw_lee@etri.re.kr

***Abstract*—A viewport adaptive low-latency streaming system for large-scale multi-view service is proposed. The proposed system is implemented based on MPEG-DASH and supports a total of 25 viewpoints. Because video segments for all 25 views cannot be transmitted at the same time due to network capacity limitations, the video segments for a main view and 8 surrounding views are transmitted. At this time, the video qualities for the main view and the surrounding views are 4K and 2K, respectively. The simulation results show the average latency of the proposed system is less than 1 msec.**

Keywords—MPEG-DASH, adaptive streaming, low latency.

I. Introduction

In recent times, large-scale multi-view streaming services (LMSS) have become increasingly important for Internet users. These services play a crucial role in various online activities such as 360-degree video streaming, real-time gaming, and video conferencing, where users expect to smoothly enjoy rich multimedia content.

However, the successful delivery of LMSS is challenged by several technical constraints. Firstly, the capacity of transmission networks is reaching its limits, imposing restrictions on the rapid transmission of massive amounts of data. Consequently, users may struggle to buffer and play back large video streams smoothly, leading to a degradation in quality of experience (QoE).

Secondly, smooth viewpoint transitions are essential in LMSS. Users should be able to transition seamlessly between various content streams, which is particularly crucial in real-time services like video conferencing or gaming. Any delays or interruptions during transitions can cause user discomfort and undermine trust in the service quality.

Lastly, low-latency transmission is of utmost importance in LMSS. Real-time applications require minimal transmission delays. For instance, real-time gaming necessitates immediate response to user inputs, while video conferencing requires effective real-time conversations. Thus, achieving low-latency transmission is essential for improving the quality of LMSS.

In order to overcome the above limitations, adaptive streaming systems have been actively researched [1] – [4].

In this paper, we propose and implement a viewport-adaptive low-latency streaming system. Our goal is to ensure smooth viewpoint transitions, and achieve low-latency transmission. We validate the effectiveness of the proposed system through simulation.

II. Viewport adaptive low-latency streaming system

A. Overview

The proposed system provides streaming service for a total of 25 views. Video images are composed of 4K, 2K and Depth encoded with AV1. The average bit rate for each image quality is shown in Table 1.

TABLE I. Bitrates each of 4K and 2K

Video quality	*Average Bitrate [Mbps]*
4K(RGB+Depth)	23 (20 + 3)
2K(RGB+Depth)	3 (2 + 1)

The client is implemented in a C++ based Windows OS environment. Assuming to stream the number of 25 of 2K videos, a network bandwidth of 100 Mbps is required. In a limited network bandwidth situation, it is impossible to stream the videos for all viewpoints. Thus, instead of receiving video segments for all viewpoint, the client downloads both 4K and 2K segments for the main view and 2K segments for the 8 surrounding views. At this time, the average bit rate of the downloaded segment is 50 Mbps. Fig. 1 shows the main view and the surrounding views when the index of the main view is 12.

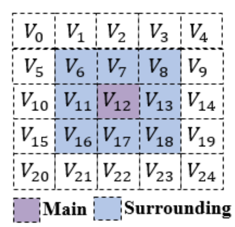
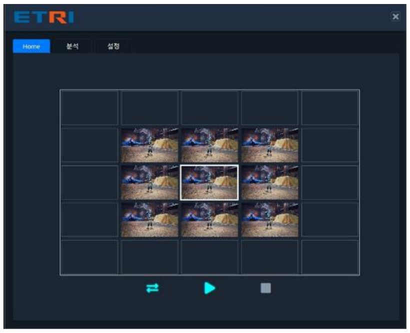


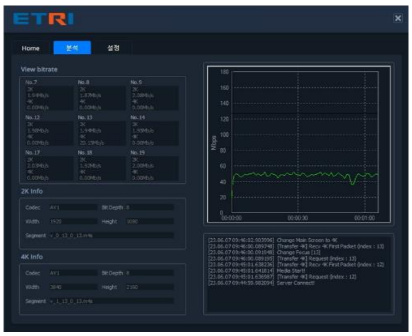
Fig. 1. Illustration of the main view and the surrounding views

The duration of the segment is 1 second. When switching views, the client plays the pre-downloaded 2K video of the corresponding viewpoint and then plays the 4K video after downloading the requested 4K segment. The shorter the duration of the segment, the less download time and view switching latency.

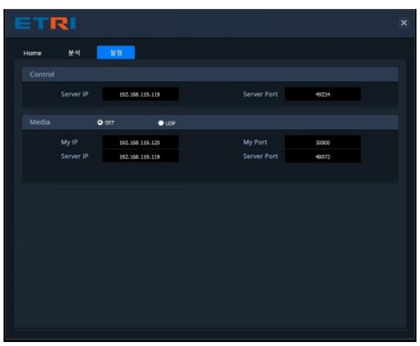
The controller of the client consists of the home, analysis, and configuration tabs. The home tab performs the functions of connecting and disconnecting from the server and controlling viewpoint switching. The Analysis tab consists of the bitrates of main view and surrounding views, 2K and 4K properties, the graph for average download throughput, and logs. The configuration tab performs the IP and port addresses of control and media channels, and transport protocol settings such as secure reliable transport (SRT) [5] and user datagram protocol (UDP). Fig. 2 shows the controller of the client described above.



(a)



(b)



(c)

Fig. 2. Controller of proposed system: (a) home tab, (b) analysis tab, (c) configuration tab

Fig. 3 shows RGB and Depth images of the main view being played.

Fig. 3. Controller of proposed system: (a) home tab, (b) analysis tab, (c) configuration tab

B. System structure

Fig. 4 shows the structure of the viewport adaptive low-latency streaming system. The proposed system is based on MPEG-DASH. MPEG-DASH is a dynamic adaptive video streaming technology based on hypertext transfer protocol (HTTP). The server transmits video divided into several segments, and the client receives and plays those segments. Because it is transmitted in segments, video quality can be dynamically adjusted to adapt to changes in network bandwidth.

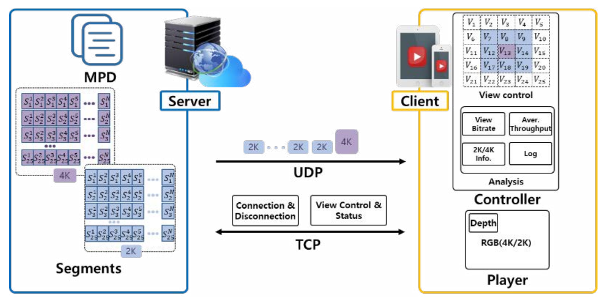


Fig. 4. Structure of the viewport adaptive low-latency streaming system

The server consists of a media presentation description (MPD) and segments, which consist of 4K and 2K segments contained RGB and depth, corresponding to 25 viewpoints per a content. The client consists of a controller and a player. Between the server and client, the control messages and segments are exchanged over transfer control protocol (TCP) and UDP, respectively.

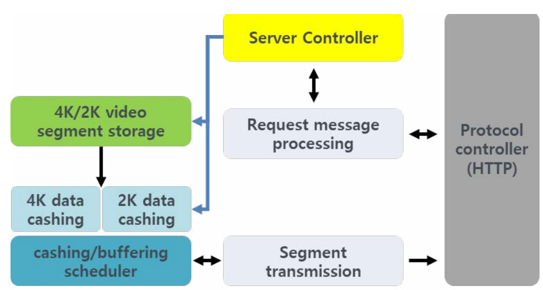


Fig. 5. Structure of the proposed server

Fig. 5 shows the structure of the proposed server. Descriptions of the main function modules of the server are as follows.

- 4K/2K video segment storage: 4K/2K video segments storage and management
- 4K/2K data cashing and buffering scheduler: 4K/2K video segments scheduling according to viewpoint switching
- Server controller: sequence and protocol control for segment transmission, and monitoring
- Request message processing: analysis of the request messages from the client, and forwards the message to the server controller

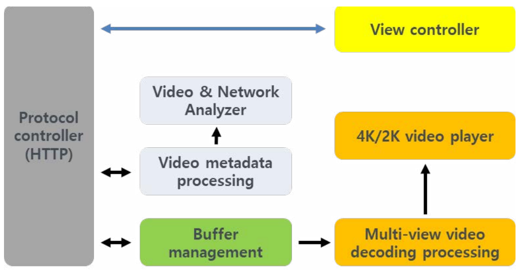


Fig. 6. Structure of the proposed client

Fig. 6 shows the structure of the proposed client. Descriptions of the main function modules of the client are as follows.

- Video & network analyzer: video and network status output
- Video metadata processing: video metadata analysis
- Buffer management: segments scheduling in playback buffer
- Multi-view video decoding processing: 4K/2K video segments decoding
- 4K/2K video player: 4K/2K RGB and depth play
- View controller: view switching control

III. Protocol Design for Viewport Adaptive Low-Latency Streaming

A. Control massage design

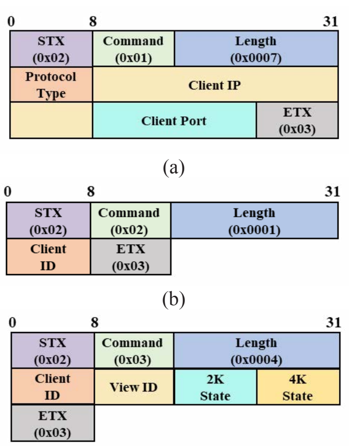


Fig. 7. Control massages: (a) client registraiton, (b) client close, (c) view switching

Fig. 7 shows the designed control message structure. The control messages between the server and client consist of command, length, and body. The start transmission (STX) 0x02 and end transmission (ETX) 0x03 are inserted to indicate the start and end of the control message. The command is defined from 0x01 to 0x04 and the contents are as follows.

- Client Registration (0x01): transmission UDP/SRT pro-tocol types, client IP, and port address to the server
- Client close (0x02): transmission client ID for client log-out to the server
- View switching (0x03): setting 4K/2K segment each viewpoint to request. The client ID, view ID, 2K, and 4K state transmission
- Main view notification (0x04): transmission the main viewpoint of client to the server. The client ID and view ID transmission

In the view switching request message, if the value of 2K and 4K state is 1, it means that the client requests segments of that view to the server, and if it is 0, it is the opposite.

B. Media transmission message design

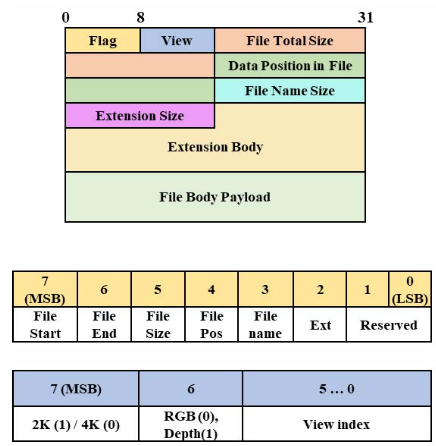


Fig. 8. Media transmission massage

Fig. 8 shows the structure of the designed media transmission message. The flag field notes the start and end of a packet and whether there are fields defined in the message such as file size, position, and name. The View field notes the video quality, type, and view index.

IV. Simulation Results

The OS environments of the server and client are centos 7.6 and windows 10, and they are all implemented based on C++. When the view switching, latency is defined as the time from when the client requests the 4K segment for that view to when the first packet of that segment is received.

Fig. 9 shows the controller screen when the main view is switched from the 12th view index to the 13th view index. Fig. 10 shows the analysis screen for the situation described in Fig. 9. The latency at this time is about 1.506 msec. In order to accurately measure, Wireshark program was used. Fig. 11 shows the packets captured through Wireshark for the time the 4K segment request for 13th view index and when the requested the first packet of that segment was received. At this time, the latency is about 1.488 msec, which is not significantly different from the value checked through the log of the analyzer. The average latency obtained by attempting 12 view switching attempts is 0.783 msec.

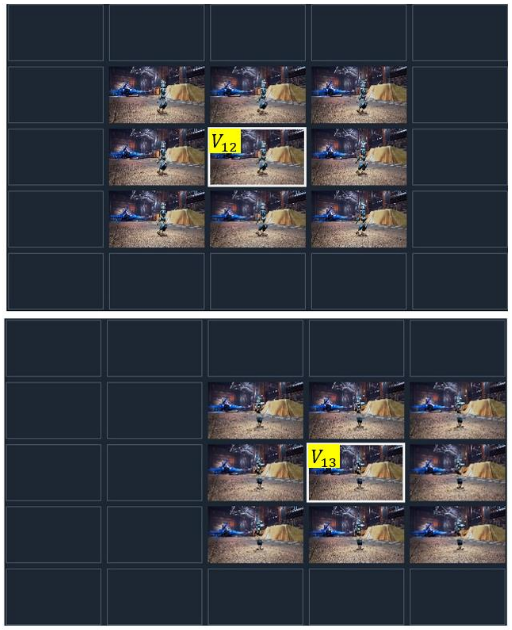


Fig. 9. Controller screen when the main view switching

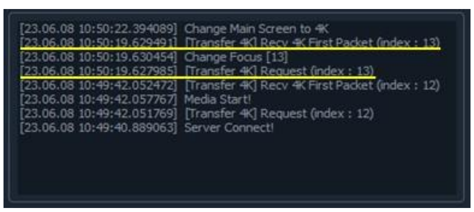
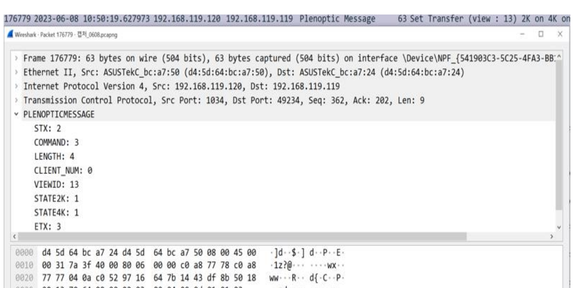
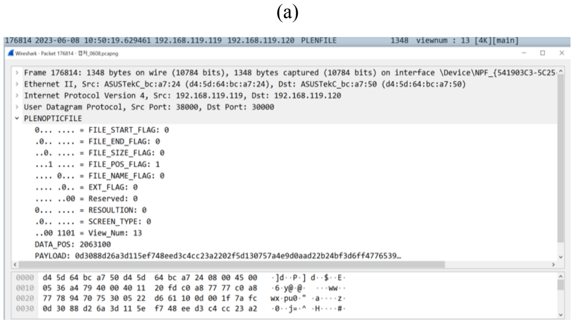


Fig. 10. Log of the analysis screen



(a)



(b)

Fig. 11. Wireshark capture screen: (a) 4K segment request, (b) the first packet for the requested 4K segment

V. Conclusion

In this paper, we developed and evaluated a viewport-adaptive low-latency streaming system for the LMSS. The proposed system utilizes a protocol designed to overcome the limitations of achieving low-latency transmission. By introducing viewport-adaptive technology, it simultaneously transmits 4K video quality of the content corresponding to the main view and 2K video quality of the surrounding views. This enables users to experience optimal video quality and smooth transitions.

The simulation results show that the proposed system demonstrates exceptionally low latency, with an average delay of less than 1 millisecond. This performance is crucial for real-time applications in various domains where low latency is required.

Further research includes field testing and user feedback. Through this, we plan to improve the performance of the proposed system and continue to develop it.

Acknowledgment

This work was supported by Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2020-0-00920 , Development of ultra high resolution unstructured plenoptic video storage/compression/streaming technology for medium to large space)

References

[1] X. Corbillon, G. Simon, A. Devlic and J. Chakareski, "Viewport-adaptive navigable 360-degree video delivery", Proc. IEEE Int. Conf. Commun., pp. 1-7, May 2017.

[2] D. Yun and K. Chung, "DASH-based multi-view video streaming system," IEEE Trans. Circuits Syst. Video Technol., vol. 28, no. 8, pp. 1974–1980, Aug. 2018.

[3] Zhang, Xue, et al. "Adaptive streaming in interactive multiview video systems", IEEE Trans. Circuits Syst. Video Technol., vol. 29, no. 4, pp. 1130-1144, Apr. 2019.

[4] Y. Chao X. Jimin, Z. Yao and M. Anlong, "Video streaming adaptation strategy for multiview navigation over DASH", IEEE Trans. Broadcast., vol. 65, no. 3, pp. 521-533, Sep. 2019.

[5] "SRT Allicance", Secure Reliable Transport Protocol, 2018, [online] Available:https://github.com/Haivision/srt/files/2489142/SRT_Protocol_TechnicalOverviewDRAFT 2018-10-17.pdf.