MultiSphere: Latency Optimized Multi-User 360° VR Telepresence with Edge-Assisted Viewport Adaptive IPv6 Multicast

Dieter Frehlich*
University of British Columbia
Vancouver, BC, Canada
frehlid@student.ubc.ca

Xincheng Huang*
University of British Columbia
Vancouver, BC, Canada
xincheng.huang@ubc.ca

Robert Xiao
University of British Columbia
Vancouver, BC, Canada
brx@cs.ubc.ca

Abstract

360° video telepresence with VR enables immersive remote collaboration, but scaling to multiple users is subject to bandwidth and latency constraints. We present *MultiSphere*, a multi-user edge-assisted 360° VR telepresence system, that combines viewport-adaptive IPv6 multicast tiling with a novel dual keyframe interval (KeyInt) streaming technique. Our approach addresses the latency bottleneck inherent in joining live streams of video using standard video codecs while maintaining visual quality through strategic use of low and high KeyInt streams. Our system achieves 75-94% bandwidth savings and an average request-to-decode latency of 56 ms, a 79% reduction compared to using a regular single-KeyInt stream.

CCS Concepts

• **Human-centered computing → Mixed / augmented reality**.

Keywords

Telepresence, 360°Video, Multicast, Edge Computing

ACM Reference Format:
Dieter Frehlich, Xincheng Huang, and Robert Xiao. 2025. MultiSphere: Latency Optimized Multi-User 360° VR Telepresence with Edge-Assisted Viewport Adaptive IPv6 Multicast. In *31st ACM Symposium on Virtual Reality Software and Technology (VRST '25), November 12–14, 2025, Montreal, QC, Canada.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3756884.3770539

1 Introduction

360° video telepresence creates immersive remote collaboration experiences with high visual fidelity, enabling applications such as virtual touring, remote instruction, and collaborative prototyping [1, 3]. Scaling to multiple users benefits group telepresence scenarios like immersive watch parties and collaborative activities.

Prior research has developed efficient 360° streaming through tile-based viewport adaptation [6] and foveated compression [2]. Researchers also proposed using multicast to save bandwidth in tile-based viewport adaptation [5]. Recent advances include multi-tier bitrate adaptation [7] and VR-specific compression techniques [4]. While these approaches optimize streaming efficiency, codec-level latency from keyframe intervals remains a fundamental bottleneck that introduces noticeable delays in tile-based approaches. We present *MultiSphere*, a multi-user 360° edge-assisted VR telepresence system targeting codec-level latency optimization, Our dual KeyInt streaming technique reduces end-to-end latency by 79% while *MultiSphere*'s IPv6 multicast delivery minimizes bandwidth consumption by 75-95%. *MultiSphere* is also backward compatible with IPv4 multicast. We believe *MultiSphere* can be valuable for future research on the group remote collaborative dynamics.

2 The MultiSphere System

MultiSphere is comprised of a commodity 360° camera, a multicasting edge-computing server, and VR clients (Fig. 1). The 360° camera produces 5760x2880 video at 30fps and sends it to the server through an Ethernet connection. The server decodes frames, splits them into a 4×8 grid of tiles, and then encodes each tile into low and high key-frame interval streams. Each stream is delivered on separate IPv6 multicast addresses, and VR clients subscribe only to the tiles which have visible pixels in the user's FoV to save bandwidth.

Dual-KeyInt Latency Optimization: H.264 keyframe intervals (KeyInt) can create significant latency when joining live streams, as clients must wait for an Instantaneous Decoder Refresh (IDR) I-frame to begin decoding. However, higher KeyInt values are crucial for maintaining compression efficiency, and reducing the interval to minimize latency noticeably harms visual quality. As such, we developed a dual KeyInt technique to address this problem by streaming two versions of every tile: a low KeyInt stream (KeyInt=3) for rapid joining, and a high KeyInt stream (KeyInt=60) for quality. Clients subscribe to both streams, display the low KeyInt version first, then switch to the high KeyInt stream upon receiving its first IDR frame and unsubscribe from the low KeyInt stream. IPv6 multicast eliminates redundant transmissions by serving multiple clients from a single stream, with brief consumption peaks only during viewport transitions (Fig. 2 (a)).

3 System Evaluation

We evaluated *MultiSphere* using a hybrid setup with 1 physical VR client and 16 simulated clients in a binary-tree topology (Fig. 3 in Appendix). We tested with 3 different KeyInt configurations: (1) a single KeyInt stream at 8 Mbps, (2) two KeyInt streams (low and high) both at 8 Mbps, and (3) two KeyInt streams (low and high) at 24 Mbps and 8 Mbps respectively. In all cases, the bitrates represent the entire 360° video distributed evenly across tiles.

We first measured bandwidth at each network node to assess the efficiency of tile-based multicast. We then measured the latency from a client sending a multicast join request to the reception of the first keyframe, capturing user perceived delay during head

*Both authors contributed equally to this research.

ACM ISBN 979-8-4007-2118-2/2025/11
https://doi.org/10.1145/3756884.3770539

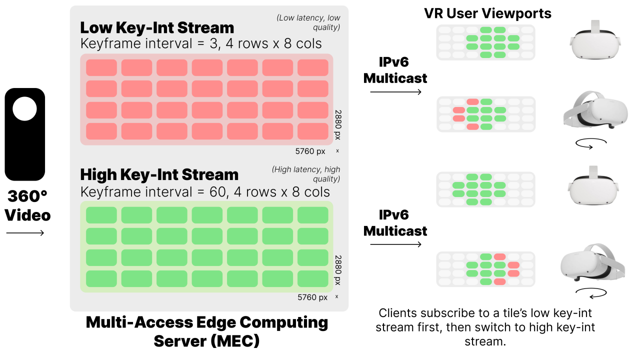

Figure 1: *MultiSphere* system architecture

movements. Each profiling run lasted 90s, during which the clients request video tiles from a pre-recorded 360° video. Clients followed synthetic head movements at 4 speeds (7.5°/s, 15°/s, 30°/s, and 60°/s) with 10s idle periods between each movement phase, using randomized starting positions to create diverse viewports.

Bandwidth Results: Compared to regular unicast streaming, our system achieved 94%, 87.5%, and 75% bandwidth reductions at the root node of the tree for single KeyInt (8Mbps), dual KeyInt (8+8Mbps), and dual KeyInt (24+8Mbps) configurations respectively (Fig. 4 in Appendix). At the client level, the tile-based viewport adaptation reduced bandwidth by 71% for single and dual-KeyInt (8+8Mbps) configurations out of the full bitrate. Fig. 2 (a) demonstrates the real-time bandwidth consumption of a single VR client.

Latency Results: Dual KeyInt streaming reduced the average tile request-to-decode latency to 56 ms, a 79% improvement over a standard single-KeyInt (KeyInt=60) stream (Fig. 2 (b)). Note that a KeyInt of 60 represents a worst case scenario; the encoder may choose a smaller KeyInt where bitrate constraints permit.

Our results in Fig. 2 demonstrate that *MultiSphere*'s dual KeyInt streaming technique provides latency comparable to a single stream with a KeyInt of 3 frames, while only consuming marginally more bandwidth during moderate head movements. This provides a balance between the superior quality of streams with a high KeyInt, and the improved request-to-decode latency of streams with a lower KeyInt without sustained bandwidth overhead.

4 Conclusion and Future Work

In this work, we present *MultiSphere*, a lightweight multi-user 360° VR telepresence system that combines IPv6 multicast with novel dual-KeyInt streaming to achieve 75-94% bandwidth savings and 79% latency reduction. Future work can 1) systematically evaluate KeyInt selection criteria under various network conditions, 2) conduct real-world deployment validation beyond simulated environments, and 3) evaluate user experience with their natural head movements in collaborative tasks our system's perceptual quality. For the VRST community, we envision *MultiSphere* as a research platform for future studies in group telepresence. For example, *MultiSphere* can support immersive group meeting or remote education in VR, fostering future research on collaborative awareness requirements and interactive techniques in such applications.

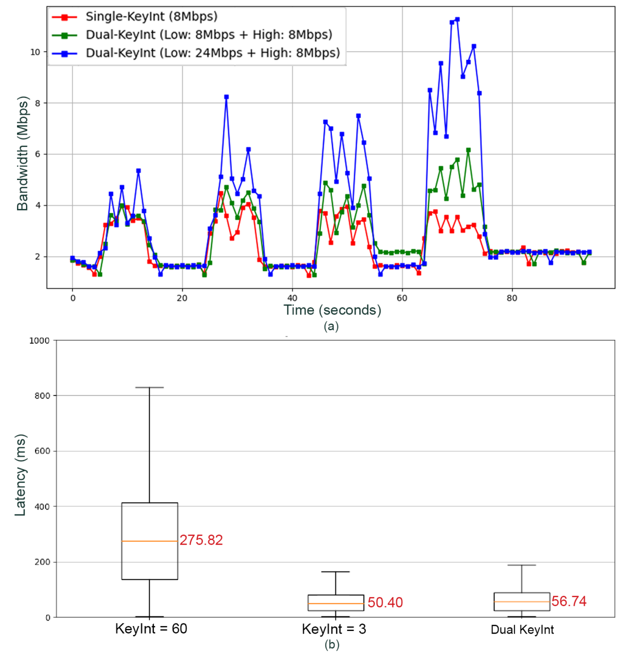

Figure 2: Bandwidth and latency evaluation results.

Acknowledgments

This work was supported by the Natural Science and Engineering Research Council of Canada (NSERC) under Discovery Grant RGPIN-2019-05624. We also thank Junkai Ding, for his help with the initial exploration on IPv6 Multicast packages.

References

[1] Xincheng Huang, Dieter Frehlich, Ziyi Xia, Peyman Gholami, and Robert Xiao. 2025. GaussianNexus: Room-Scale Real-Time AR/VR Telepresence with Gaussian Splatting. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology (UIST '25)*. Association for Computing Machinery, New York, NY, USA, Article 189, 18 pages. doi:10.1145/3746059.3747693

[2] Xincheng Huang, James Riddell, and Robert Xiao. 2023. Virtual Reality Telepresence: 360-Degree Video Streaming with Edge-Compute Assisted Static Foveated Compression. *IEEE Transactions on Visualization and Computer Graphics* 29, 11 (2023), 4525–4534. doi:10.1109/TVCG.2023.3320255

[3] Xincheng Huang, Michael Yin, Ziyi Xia, and Robert Xiao. 2024. VirtualNexus: Enhancing 360-Degree Video AR/VR Collaboration with Environment Cutouts and Virtual Replicas. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology* (Pittsburgh, PA, USA) *(UIST '24)*. Association for Computing Machinery, New York, NY, USA, Article 55, 12 pages. doi:10.1145/3654777.3676377

[4] Luoyu Mei, Shuai Wang, Yun Cheng, Ruofeng Liu, Zhimeng Yin, Wenchao Jiang, Shuai Wang, and Wei Gong. 2024. ESP-PCT: Enhanced VR Semantic Performance through Efficient Compression of Temporal and Spatial Redundancies in Point Cloud Transformers. In *Proceedings of the Thirty-ThirdInternational Joint Conference on Artificial Intelligence (IJCAI-2024)*. International Joint Conferences on Artificial Intelligence Organization, 1182–1190. doi:10.24963/ijcai.2024/131

[5] Tsubasa Okamoto, Takumasa Ishioka, Tatsuya Fukui, Ryouhei Tsugami, Toshihito Fujiwara, Satoshi Narikawa, Takuya Fujihashi, Shunsuke Saruwatari, and Takashi Watanabe. 2024. End-to-End Multi-User 360-Degree Video Delivery Using Users' Fixation Points. *IEEE Access* 12 (2024), 9151–9161.

[6] Feng Qian, Bo Han, Qingyang Xiao, and Vijay Gopalakrishnan. 2018. Flare: Practical Viewport-Adaptive 360-Degree Video Streaming for Mobile Devices. In *Proceedings of the 24th Annual International Conference on Mobile Computing and Networki ng* (New Delhi, India) *(MobiCom '18)*. Association for Computing Machinery, New York, NY, USA, 99–114.

[7] Farzad Tashtarian, Mahdi Dolati, Daniele Lorenzi, Mojtaba Mozhganfar, Sergey Gorinsky, Ahmad Khonsari, Christian Timmerer, and Hermann Hellwagner. 2025. ALPHAS: Adaptive Bitrate Ladder Optimization for Multi-Live Video Streaming. In *IEEE INFOCOM 2025 - IEEE Conference on Computer Communications*. 1–10.