Innovating Multi-user Volumetric Video Streaming through Cross-layer Design

Ding Zhang
George Mason University
dzhang13@gmu.edu

Bo Han
George Mason University
bohan@gmu.edu

Parth Pathak
George Mason University
phpathak@gmu.edu

Haoliang Wang
Adobe Research
hawang@adobe.com

ABSTRACT

Although existing work has demonstrated the feasibility of streaming volumetric content to a single user, there exist many appealing applications (*e.g.,* classroom education and collaborative design) that involve multiple users who watch the same volumetric content simultaneously. In this paper, we first perform a scaling experiment to demonstrate the challenges of streaming high-quality volumetric videos to multiple users and reveal the viewport-similarity opportunity that we can leverage to effectively optimize the network resource utilization using multicast over mmWave. We then develop a holistic research agenda for improving the performance and quality of experience for multi-user volumetric video streaming on commodity devices. Our proposed research includes joint viewport prediction and blockage mitigation for multiple users, multicast grouping based on viewport similarity, customized mmWave beam design for efficient multicast, and mmWave-aware multi-user video rate adaptation. Finally, we discuss the open challenges of building a practical system with the proposed research roadmap.

ACM Reference Format:
Ding Zhang, Bo Han, Parth Pathak, and Haoliang Wang. 2021. Innovating Multi-user Volumetric Video Streaming through Cross-layer Design. In *The Twentieth ACM Workshop on Hot Topics in Networks (HotNets '21), November 10–12, 2021, Virtual Event, United Kingdom.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3484266.3487396

1 INTRODUCTION

Volumetric video is an emerging 3D video format that enables users to explore multimedia content with six degrees of freedom (6DoF) motion, 3DoF of translational movement (X, Y, and Z), and another 3DoF of rotational movement (yaw, pitch, and roll). Different from the traditional 2D videos that are based on a fixed view and the 360-degree videos that support only the 3DoF rotational movement, volumetric videos provide a more interactive and immersive experience for viewers. When used in augmented reality (AR) or virtual reality (VR), volumetric videos can enable many promising applications in entertainment, education, healthcare, *etc.* For example, users can track and view every movement of their favorite players during a sports event by moving around with 6DoF. Telepresence with volumetric videos can help students attend classes remotely with an immersive learning experience. Therefore, it has been regarded as the key application in 5G and beyond [16, 18].

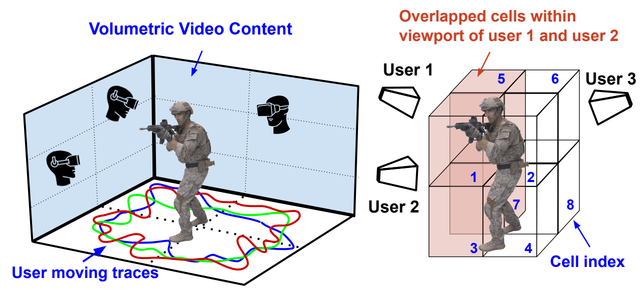


Figure 1: Multi-user volumetric video streaming with overlapped cells.

Although many of the applications of volumetric videos are specific to multi-user scenarios, limited attention has been paid to multi-user volumetric video streaming. The main reason is that volumetric video streaming is bandwidth-hungry and computation-intensive. Thus, streaming to a single user is already a challenging problem. Among the state-of-the-art volumetric video streaming systems, ViVo [6] explores viewport, distance, and occlusion optimizations to save the bandwidth. As a result, it demands a 100 to 200 Mbps data rate for a single user when watching point-cloud video frames with around 200K points. Transmitting high quality (high point density) frames and more complex scenes may make the bandwidth requirement even higher. GROOT [12] also requires 100 to 500 Mbps bandwidth for a single user after applying novel compression and GPU-assisted decoding schemes to reduce the computation overhead. However, in the multi-user scenario, the total consumed bandwidth

HotNets'21, November 10-12, 2021, Virtual Event, UK

ACM ISBN 978-1-4503-9087-3/21/11…$15.00
https://doi.org/10.1145/3484266.3487396

increases linearly with the number of users, requiring a few Gbps bandwidth from the network.

In this paper, we design a multi-user volumetric video system over mmWave WLANs, by exploring the offered high data rate. We show that while mmWave WLANs can provide sufficient bandwidth for supporting three to four users watching volumetric videos simultaneously, we need to leverage multicast and cross-layer design to offer high-quality content and scale the system for more users. This paper explores multiple research directions in these directions. We list the challenges we are facing and our contributions below.

- There is limited understanding of the performance of multi-user volumetric video streaming using WLANs. Thus, we measure the multi-user performance of state-of-the-art system ViVo [6] with both 802.11ac and 802.11ad WLANs. We find that 1) 802.11ac WLANs cannot support multi-user volumetric video streaming even for low-quality videos; 2) 802.11ad mmWave WLANs can support a limited number of users but still need to be further optimized to support more users; 3) users share high viewport similarity when watching volumetric videos, which can be utilized in the multicast transmission to optimize the bandwidth requirement.
- In multi-user scenarios, the traditional viewport prediction methods cannot be directly used as the users might occlude the viewport of each other and affect the moving directions. We aim to build a model for joint viewport prediction for multiple users by incorporating their interactions. In addition, this multi-user viewport prediction could be used to estimate the blockage events in mmWave WLANs, which can enable a fast, cross-layer, proactive blockage mitigation.
- Building an efficient multicast system to stream volumetric videos is challenging, and the directional communication in mmWave can lead to unbalanced received signal strength (RSS) that makes the multicast transmission over mmWave even harder. We design a multicast scheme that selects user groups with high viewport similarity to reduce the transmission latency. Moreover, we employ a new custom beam design scheme to overcome the unbalanced RSS problem. Our preliminary results show that the new beam design can effectively improve the multicast transmission performance in mmWave WLANs.
- To adapt to the throughput fluctuations in mmWave WLANs, we design a cross-layer bandwidth prediction scheme by combining the data rate indicators from the physical layer (blockage or mobility) and the application layer (buffer size or throughput). The bandwidth prediction can guide the multicast scheduler to perform judicious actions (prefetching, video quality adaptation, or beam switching) and offer high quality of experience (QoE) for users.

We are currently working on realizing the above ideas and integrating them into a holistic multi-user volumetric video streaming system over mmWave WLANs.

2 RELATED WORK

Multi-user AR, VR, and 360° Video Streaming. Recent work has started to address the technical issues of supporting multiple users for AR [22, 36], VR [13–15], and 360° video streaming [3, 4, 20]. For example, SPAR [22] aims to reduce the spatial inconsistency of visual content and high initialization latency in multi-user AR scenarios by adapting the communicated information to the positions of virtual objects. Coterie [15] enables multi-user VR on commodity mobile devices by exploiting the similarity of background content in consecutive frames of the same user to reduce the bandwidth consumption. Based on a simulation study, Bao *et al.* [3] demonstrated the effectiveness of using multicast to transmit the shared content in 360° videos to multiple users. To the best of our knowledge, we are the first to investigate the research challenges of multi-user volumetric video streaming by jointly considering the unique features of volumetric content and mmWave networks.

Volumetric Video Streaming. Existing work on volumetric video streaming has been centered around the single-user scenario [6, 12, 19, 21, 34]. For example, ViVo [6] utilizes viewport, distance, and occlusion optimizations to reduce the bandwidth requirement of mobile volumetric video streaming. GROOT [12] presents a GPU-assisted point cloud compression scheme to optimize the decoding latency of volumetric content on mobile devices. VoluSR [34] is a recent proposal that leverages 3D super-resolution to upsample point clouds for improving user experience. Different from the above work, we focus on multi-user volumetric video streaming and propose a research agenda that benefits from a cross-layer design by jointly optimizing mmWave networks and volumetric content delivery.

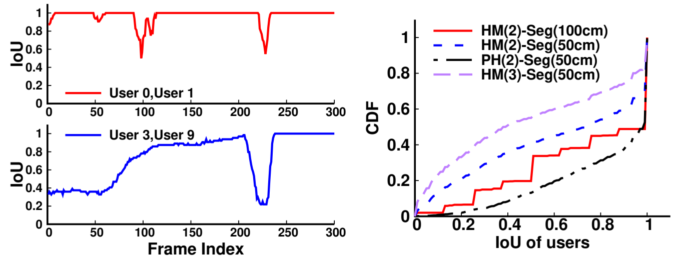
mmWave WLANs. Blockage and mobility are two key technical challenges in mmWave networks. A variety of approaches have been proposed to tackle these problems at the mmWave physical/MAC layer [8, 24, 25, 28, 29, 35]. Moreover, mmWave links have been leveraged to replace the cables that connect high-end tethered VR devices and powerful graphics-rendering machines, by designing a mirror-like transceiver to address the blockage problem [1]. In terms of multicast over mmWave, Naribole *et al.* [17] proved its feasibility for highly directional 60-GHz WLANs. In contrast, we aim to address the practical challenges when utilizing mmWave multicast for bandwidth-hungry multi-user volumetric video streaming through a cross-layer design that takes advantage of the plethora of existing work.

3 MOTIVATION

In this section, we first identify the challenges of supporting multiple users in volumetric video streaming, by experimentally evaluating a prototype that extends ViVo [6], the

	Num. of Users	Per user data rate (Mbps)	FPS: Vanilla 330K points	FPS: Vanilla 430K points	FPS: Vanilla 550K points	FPS: Multi-user ViVo 330K points	FPS: Multi-user ViVo 430K points	FPS: Multi-user ViVo 550K points
ac	1	374	30	30	30	30	30	30
	2	180	21.5	17.4	14.1	30	28.5	21.9
	3	112	13.6	10.9	8.4	19.2	17.7	13.6
ad	1	1270	30	30	30	30	30	30
	2	575	30	30	30	30	30	30
	3	382	30	30	30	30	30	30
	4	298	30	29.3	21.8	30	30	30
	5	231	27.4	21.6	18.0	30	30	29.3
	6	175	19.8	16.5	13.2	30	27.5	21.2
	7	144	16.8	13.5	11.2	27.0	22.9	17.2

Table 1: Performance of multi-user volumetric video streaming with *vanilla* and ViVo [6] systems.



(a) (b)

Figure 2: Viewport similarity in multi-user volumetric video streaming (a) over time, and (b) for different device types, different partition sizes, and different group sizes.

state-of-the-art volumetric video streaming system, for the multi-user scenario. We then analyze the 6DoF viewport trajectories from an IRB-approved user study that reveal significant overlaps of volumetric content consumed by different users, which motivates our proposed research agenda.

Experimental setup. Our testbed consists of a content server that hosts multiple volumetric videos. The server is a machine running Ubuntu 18.04 and it also acts as an access point (AP). It has both 802.11ac and 802.11ad wireless interfaces. We use multiple laptops as the clients, which are equipped with Intel i7 CPU (4 cores at 2.8GHz) and 8GB RAM. This configuration is comparable to state-of-the-art smartphones such as Samsung Galaxy S21 and is slightly better than the Magic Leap One mixed reality headset, which uses the NVIDIA Jetson TX2 hardware. We could not use these mobile devices directly for our experiments because they do not have an 802.11ad interface. Both the server and the clients have an 802.11ad card with Qualcomm QCA9500 chipset supported by wil6210 [30] driver.

We use the soldier volumetric video released in the 8i dynamic voxelized point cloud dataset [11] as our video source. This video was captured at 30 FPS. Due to the huge size of point cloud data, we use the Draco library [5] from Google to compress this video and reduce its bandwidth consumption for streaming. We create three versions of this video with different visual qualities, by varying the numbers of points per frame. The version with the highest quality has, on average, 550K points per frame, which is the highest point density that can be decompressed by Draco at 30 FPS on the client laptops. Note that this point density doubles the highest density of the volumetric videos used in ViVo [6]. The medium and low qualities have 430K points/frame and 330K points/frames, respectively. After the compression, the bitrate of these different versions ranges from 235 to 364 Mbps. We implement two video players on the client laptops. One is the vanilla system that fetches the entire point cloud for each video frame, and the other implements the viewport, occlusion, and distance optimizations in ViVo for reducing the data usage. Next, we benchmark both the vanilla system and ViVo with the above three visibility-aware optimizations when they are used for supporting multiple users, by measuring the maximum achievable frame rate.

Unicast over 802.11ac and 802.11ad for multiple users. To examine the multi-user volumetric video streaming performance, we deliver the same video content [11] from the server to different numbers of concurrent users over either an 802.11ac or 802.11ad network with unicast. On our testbed, when serving a single user, the throughput is around 374 Mbps for 802.11ac and 1270 Mbps for 802.11ad mmWave network. Table 1 shows the maximum achievable frame rate (capped at 30 FPS) for different numbers of users when streaming the volumetric video with different point densities.

We have the following key observations from Table 1. First, when using 802.11ac, the vanilla system can support only one user at 30 FPS, regardless of the video quality. When multiple users watch the volumetric video simultaneously over the same 802.11ac network, the maximum frame rate drops below 30 FPS, which deteriorates the user experience. ViVo can help only for the low-quality video, by adding one more user. Second, as the mmWave 802.11ad network at 60 GHz can achieve much higher throughput than 802.11ac (*e.g.*, higher than 1 Gbps without blockage), it can indeed increase the number of concurrent users to three for the medium and high quality videos and four for the low-quality video. ViVo can further increase the number of users by one or two, depending on the video quality. Third, the maximum number of users that can be supported by ViVo at 30 FPS is only four for the high-quality video. However, use cases such as group-watching of sports events or AR-enhanced classroom teaching may involve more users. This limited number of users that can simultaneously watch high-quality volumetric content over an 802.11ac/802.11ad network with the state-of-the-art system motivates us to explore the opportunities for improving the performance and QoE of multi-user volumetric video streaming.

Inter-user viewport similarity. Viewport similarity has been leveraged for tile-based multi-user 360° video streaming, for which the tiles in the overlapped viewport of multiple users will be delivered using multicast [3]. However, given the 6DoF movement that is enabled by volumetric videos, it is unclear whether there still exists significant viewport similarity when multiple users consume the same volumetric video in the 3D space. Thus, we investigated the viewport trajectories collected from 32 participants when watching several volumetric videos in an IRB-approved user study [6]. The 6DoF viewport trajectories were collected for all users at 30 Hz during the viewing sessions. The participants were divided into two groups, one using a smartphone and the other using a Magic Leap One headset to watch the volumetric videos. We denote these two groups as PH and HM, respectively.

To calculate the 3D viewport similarity for volumetric video streaming between different users, we spatially partition the original point cloud into smaller "sub" regions called *cells*, which has been used in viewport-adaptive volumetric video streaming systems such as ViVo [6]. Each cell is independently prefetchable and decodable. We partition the entire point cloud into cells of three sizes: 25×25×25 cm^3, 50×50×50 cm^3, and 100×100×100 cm^3. After the partition, we use frustum culling [26] to determine the cells overlapping with the 3D viewport based on the position and orientation of the user and calculate a *visibility map* that records the visible cells for each user. We then define the *viewport similarity* of a group of users as the intersection over union (IoU) of their visibility maps. For example, as shown in Fig. 1, if we segment the video content into 8 cells, cells 1, 3, and 5-8 are visible for User 1, and cells 1-4, 5, and 7 are visible for User 2. Thus, cells 1, 3, 5, and 7 will be needed for both users and their viewport similarity (IoU) for this frame is 0.5.

We analyze the viewport similarity among all 32 participants and make the following observations. First, there is a significant viewport overlap between users, which means there are indeed opportunities for using multicast to reduce the required bandwidth of multi-user volumetric video streaming. Fig. 2a shows how the IoU changes when two randomly selected users watch the same video (partitioned into 50×50×50 cm^3 cells). As we can see from this figure, User 0 and User 1 watch exactly the same content most of the time. For User 3 and User 9, although the viewport similarity is low initially, it increases to 1 towards the end of the video.

Second, viewport similarity is affected by many factors including the device type, the segmentation granularity, and the group size. Fig. 2b shows the CDF of the viewport similarity (IoU) among all users for different settings. With a headset (HM(2)-Seg(50cm)), there is less viewport similarity among users because they can move relatively more freely compared to watching the video on smartphones, which is shown by the PH(2)-Seg(50cm) curve. In terms of segmentation granularity (HM(2)-Seg(50cm) vs HM(2)-Seg(100cm)), the IoU between users decreases when the number of cells increases with finer-grained segmentation. Furthermore, the number of users that we consider also affects the viewport similarity. As shown in Fig. 2b, when calculating the viewport similarity for more users (HM(3)) in a group, the IoU decreases as more users bring larger variations in their positions and orientations.

Motivated by the above preliminary results of multi-user volumetric video streaming on 802.11ad mmWave networks and the significant viewport similarity among users, we aim to answer the following question in this paper: how to reduce the bandwidth requirement of multi-user volumetric video streaming by exploring mmWave multicast through cross-layer design? As we can see from Table 1, the bandwidth reduction can either lead to more concurrent users or improve the QoE for a given number of users.

4 RESEARCH AGENDA

4.1 Multi-user viewport prediction

Joint viewport prediction for multi-user. Exploiting viewport similarity for multicast requires us to predict the viewport for all users in advance. Existing works such as [6, 31] have shown that individual users' 6DoF can be predicted using linear regression or multilayer perceptron with high accuracy in real-time. In multi-user scenarios, to predict the viewport of all users together, it is intuitive to just combine each of them and create a holistic view of viewport mobility of all users.

However, depending on volumetric video usage in AR or VR, it is possible that one user's movement may affect the viewport of other users. For example, when watching volumetric videos in AR, one user can occlude the view of another user. In such cases, it is necessary to adapt the existing models of viewport prediction to incorporate occlusions. We will conduct more user studies for developing joint multi-user viewport prediction models by considering the user interaction. It will facilitate the mmWave channel estimation under blockages/mobility and the multicast transmission that leverages viewport similarity.

Viewport prediction for proactive blockage mitigation. Human blockages can significantly reduce the data rate of mmWave links and can even cause a complete outage of the links. Reinitiating beam searching to find new beams to go across the blockages will cause a delay of up to 5 to 20 ms. This added latency can cause video stalls and a reduction in video quality that adversely affects users' QoE. Prior works such as [28, 32] have proposed to use in-built sensors to predict user's mobility and adapt mmWave beams accordingly. However, in our multi-user scenario, one user can block the mmWave link of another user and it is not sufficient to adapt the mmWave beams to each user separately.

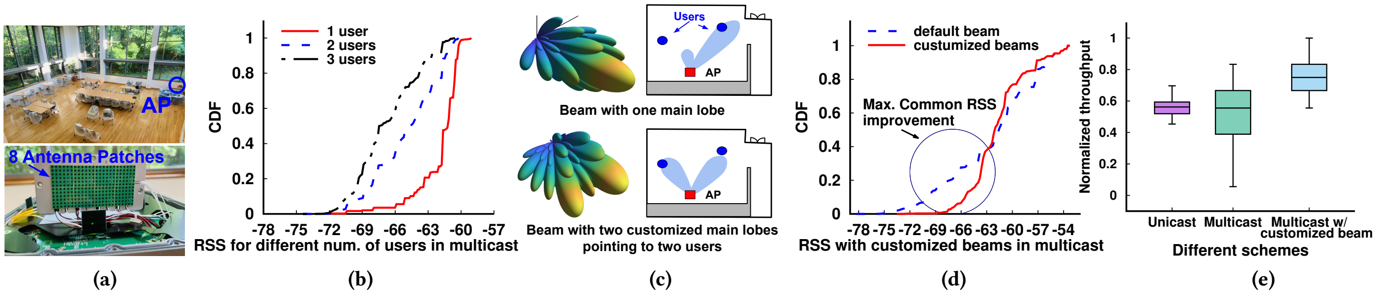


Figure 3: (a) mmWave measurement setup (b) The default beams cannot support an efficient multicast for multiple users. (c-d) With the customized beam design in multicast, we can efficiently transmit the overlapped cells between users. (e) Preliminary performance of the multicast system with customized beams.

We will leverage our multi-user viewport prediction to build a cross-layer proactive beam adaptation and blockage mitigation scheme. The holistic view of the multi-user viewport prediction available at the AP will be used to infer possible blockages between users. Afterward, the AP can estimate the data rate change under the potential blockages for each link and can take one of many possible actions at different layers proactively. It can prefetch the content and schedule the future cells in the current time slot so that when the blockage happens, it has already prefetched some frames for rendering. In addition, the AP can adapt its beam to the user with a reflection path without beam searching. Therefore, the link blockage can be effectively mitigated.

4.2 Multicast with customized beam design

Multicast grouping with viewport similarity. Based on the multi-user viewport prediction, we utilize multicast to transmit overlapped content among users to reduce the motion-to-photon latency. Specifically, we estimate the transmission time of a frame to a user group $\mathbf{k}$ as $T^m(\mathbf{k}) = S^m_{\mathbf{k}}/r^m + \sum_i^{|\mathbf{k}|}(S_i - S^m_{\mathbf{k}})/r_i$ where $S^m_{\mathbf{k}}$ is the size of the overlapped cells of user group $\mathbf{k}$, and r^m is the multicast data rate. S_i and r_i are the requested frame size and unicast data rate for user i, respectively. The first part of the equation is the time to transmit the overlapped cells with multicast while the second part is the transmission time to stream the remaining requested cells with default unicast among $|\mathbf{k}|$ users. We find the group $\mathbf{k}$ among users with high viewport similarity to satisfy the constraint of $T^m(\mathbf{k}) <= 1/F$ where the F is the required frame rate. Specifically, the throughput improvement made possible by multicast can be used to increase the transmitted frame size S_i (*i.e.*, improve the video quality) or to serve more users watching the volumetric content.

Multicast with multi-lobe beams. A key challenge with multicast is that the multicast data rate r^m for a group is limited by the lowest achievable modulation coding scheme (MCS) among all users of the group. Using this MCS guarantees a reliable multicast received by all users. In mmWave WLANs, it is challenging to cover multiple users of a group using directional beams. If sufficiently high RSS cannot be guaranteed for all members of the group, multicast may provide even worse performance than unicast.

To understand the performance of multicast in mmWave networks, we first investigate the effectiveness of the default beam codebook used in the commercial 802.11ad devices. We use the viewport traces for multiple users (Section 3) and measure the RSS for these users in our mmWave WLAN testbed. Our mmWave testbed includes an 802.11ad router from Airfide [2] with 8 phased antenna array patches (shown in Fig. 3a) and multiple Acer laptops with 802.11ad interfaces. We modify the open-source 802.11ad driver on the laptop to extract the RSS, MCS, and beamforming information in userspace. Fig. 3b shows the CDF of maximum RSS that could be supported by the default codebook for multicast groups of different sizes. We find that using the default codebook, RSS of -68 dBm, which can provide approximately 384 Mbps data rate necessary for 550K points quality, can be provided for 96.5% positions for one user (no multicast), but only 79% and 60% positions for two and three users, respectively, with multicast. This is because the default codebook beams are not specifically designed to support multicast and cannot guarantee high RSS for all members in a multicast group.

To address the problem of poor and unbalanced RSS in multicast provided by the default codebook, we design customized multi-lobe beams that can provide high RSS to all users of a given multicast group. The main idea is to create customized beams by combining the antenna weight vectors of beams pointing to individual users while constraining the total transmission power. Assume the antenna weight vector used for transmitting to User 1 is $\mathbf{w}_1$ and RSS is Δ_1, and $\mathbf{w}_2$ and Δ_2 are the weight vector and RSS for User 2, we define the combined antenna weight vector for the new two-lobe beam (shown in Fig. 3c) as $\mathbf{w} = \frac{1}{\Delta_1+\Delta_2}(\Delta_2\mathbf{w}_1 + \Delta_1\mathbf{w}_2)$. The objective here is to design beams that can not only cover the users of a group with high viewport similarity but also provide higher common MCS supported by all members. We note that we can use the predicted 6DoF motion information at the server to select the individual beams and combined beams for the AP without beam searching. Moreover, our custom beam design just requires the RSS value to normalize the antenna weights, instead of the complete channel state information (CSI) as in [8, 27] because the separated users have independent receive chain.

To investigate the effectiveness of the proposed multi-lobe beam design, we run the multicast for two users with our custom beams and default beams in a commercial mmWave channel simulator Remcom [23]. As shown in Fig. 3d, by using the new combined weight, the two users can achieve much higher common RSS values (in the circle), leading to higher common MCS. We also notice that when both users have high RSS, we should directly use the default common beam that has already covered the users.

Lastly, we compare the unicast, multicast with default beams, and multicast with custom multi-lobe beams in terms of normalized throughput for the case with two users (Fig. 3e). We find that multicast with the default beams cannot always improve the data rate but may in fact sometimes reduce the data rate. This is due to the unbalanced RSS that reduces the common MCS supported by the multicast group. However, using multicast with our customized beams can effectively increase the data rate and the performance gains can be higher with more users in the multicast group.

4.3 Cross-layer video rate adaptation

To offer better QoE for multi-user volumetric video streaming, we propose a cross-layer rate adaption scheme. We identify two important issues when exploring the design space. First, how to accurately estimate the link bandwidth (the data rate r_i and r^m) for unicast and multicast transmissions? In the multi-user scenario, an inaccurate bandwidth estimation will affect the QoE for all users in the same multicast group. To better adapt to network bandwidth changes, we aim to utilize a cross-layer solution that combines the mmWave channel information (*e.g.,* RSS) with the application layer information such as the buffer size of the video player [7]. Second, how to judiciously react to the bandwidth fluctuation of mmWave links? The possible reactions include: 1) prefetching more cells for a priority group of users with low predicted bandwidth; 2) regrouping the multicast and unicast transmissions to better utilize viewport overlaps and save more bandwidth; 3) generating a new beam to cover a different physical path to improve the channel quality. Our rate adaptation scheme will intelligently select different solutions according to the mmWave link status.

5 OPEN CHALLENGES

Viewport prediction for blockage mitigation. One key challenge of using multi-user viewport prediction to predict the blockage and throughput reduction is that blockage does not always cause link outage [25]. This requires developing a model to quantify the degradation of the link performance with respect to different levels of blockage and integrating it into the multicast scheduler.

Multicast with customized beams. Even though we have shown the effectiveness of multicast with customized beams, there are a few challenges in the design on COTS devices. First, multi-lobe beams can cause interference between users of the same group due to reflections from nearby objects. This means that we need to probe the newly designed beam to verify that it will not cause too much interference before using it. Second, the imperfections in the antenna hardware of COTS devices produce irregular patterns that can reduce the effectiveness of custom beams. Lastly, the MAC layer multicast implementation and multi-lobe beam design need heavy engineering of the driver and the firmware.

Cross-layer video rate adaptation. In contrast to traditional video rate adaptation algorithms that are executed by the players on the client side [9, 33], our proposed scheme for multi-user volumetric video streaming is realized at a central point (*e.g.,* on an edge server). A key issue is that the cross-layer approach to choose the proper video encoding bitrates for multiple users leads to a much larger design space when formulating the optimization problem, compared to the conventional application-layer schemes that mainly focus on a single user.

Multiple APs Coordination. To allow even more users to watch volumetric content at the same time, there are opportunities to utilize multiple APs, each of which can serve a specific multicast group separately. Thanks to the directional nature of mmWave links, multiple APs could serve different groups of clients concurrently to achieve high spatial reuse [10, 35]. Expanding our system to multiple APs poses other challenges including multicast grouping for different APs, interference management between multi-lobe beams, and control overhead associated with achieving the coordination.

6 CONCLUSION

In this position paper, we first show the challenges of using current 802.11 WLANs to simultaneously stream volumetric videos to multiple co-located users. Motivated by the 3D viewport similarity that we identify from a trajectory trace with 32 users, we then propose a holistic research agenda by leveraging cross-layer design to improve the performance and QoE of multi-user volumetric video streaming over customized mmWave multicast. We conclude the paper with several remaining open challenges for building a practical system, which we hope could stimulate more discussion and research on this topic.

ACKNOWLEDGMENTS

We would like to thank the anonymous reviewers for their valuable comments and feedback. This research is supported by NSF grants CNS-1815945 and CNS-1730083.

REFERENCES

[1] O. Abari, D. Bharadia, A. Duffield, and D. Katabi. 2017. Enabling High-Quality Untethered Virtual Reality. In *14th USENIX Symposium on Networked Systems Design and Implementation (NSDI 17)*. 531–544.

[2] Airfide 2020. (2020). https://airfidenet.com/

[3] Y. Bao, T. Zhang, A. Pande, H. Wu, and X Liu. 2017. Motion-Prediction-Based Multicast for 360-Degree Video Transmissions. In *2017 14th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*. 1–9.

[4] J. Chen, X. Qin, G. Zhu, B. Ji, and B. Li. 2021. Motion-Prediction-based Wireless Scheduling for Multi-User Panoramic Video Streaming. In *IEEE INFOCOM 2021*. 1–10.

[5] Google Draco library 2020. (2020). https://github.com/google/draco

[6] B. Han, Y. Liu, and F. Qian. 2020. ViVo: Visibility-Aware Mobile Volumetric Video Streaming. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking (MobiCom '20)*.

[7] T. Huang, R. Johari, N. McKeown, M. Trunnell, and M. Watson. 2014. A Buffer-Based Approach to Rate Adaptation: Evidence from a Large Video Streaming Service. In *Proceedings of the 2014 ACM Conference on SIGCOMM (SIGCOMM '14)*. 187–198.

[8] I. K. Jain, R. Subbaraman, and D. Bharadia. 2021. Two beams are better than one: Enabling reliable and high throughput mmWave links. (2021). arXiv:eess.SP/2101.04249

[9] J. Jiang, V. Sekar, and H. Zhang. 2012. Improving Fairness, Efficiency, and Stability in HTTP-Based Adaptive Video Streaming with FESTIVE *(CoNEXT '12)*. 97–108.

[10] S. Jog, J. Wang, J. Guan, T. Moon, H. Hassanieh, and R. R. Choudhury. 2019. Many-to-Many Beam Alignment in Millimeter Wave Networks. In *16th USENIX Symposium on Networked Systems Design and Implementation (NSDI 19)*. 783–800.

[11] JPEG Pleno Database: 8i Voxelized Full Bodies - A Dynamic Voxelized Point Cloud Dataset 2019. (2019). http://plenodb.jpeg.org/pc/8ilabs

[12] K. Lee, J. Yi, Y. Lee, S. Choi, and Y. Kim. 2020. GROOT: A Real-Time Streaming System of High-Fidelity Volumetric Videos. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking (MobiCom '20)*. Article 57, 14 pages.

[13] Y. Li and W. Gao. 2018. MUVR: Supporting Multi-User Mobile Virtual Reality with Resource Constrained Edge Cloud. In *2018 IEEE/ACM Symposium on Edge Computing (SEC)*. 1–16.

[14] X. Liu, C. Vlachou, F. Qian, C. Wang, and K. Kim. 2020. Firefly: Untethered Multi-user VR for Commodity Mobile Devices. In *2020 USENIX Annual Technical Conference (USENIX ATC 20)*. 165–171.

[15] Jiayi Meng, Sibendu Paul, and Y. Charlie Hu. 2020. Coterie: Exploiting Frame Similarity to Enable High-Quality Multiplayer VR on Commodity Mobile Devices. In *Proceedings of the Twenty-Fifth International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS '20)*. 923–937.

[16] A. Narayanan, E. Ramadan, J. Carpenter, Q. Liu, Y. Liu, F. Qian, and Z. Zhang. 2020. A First Look at Commercial 5G Performance on Smartphones. In *Proceedings of The Web Conference 2020 (WWW '20)*. 894–905.

[17] Sharan Naribole, Edward Knightly, Sharan Naribole, and Edward Knightly. 2017. Scalable Multicast in Highly-Directional 60-GHz WLANs. *IEEE/ACM Trans. Netw.* 25, 5 (Oct. 2017), 2844–2857.

[18] Next generation display media in 6G 2021. (2021). https://research.samsung.com/next-generation-display-media

[19] Jounsup Park, Philip A. Chou, and Jenq-Neng Hwang. 2019. Rate-Utility Optimized Streaming of Volumetric Media for Augmented Reality. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems* 9, 1 (2019), 149–162.

[20] C. Perfecto, M. S. Elbamby, J. Del Ser, and M. Bennis. 2020. Taming the Latency in Multi-User VR 360°: A QoE-Aware Deep Learning-Aided Multicast Framework. *IEEE Transactions on Communications* 68, 4 (2020), 2491–2508.

[21] Feng Qian, Bo Han, Jarrell Pair, and Vijay Gopalakrishnan. 2019. Toward Practical Volumetric Video Streaming on Commodity Smartphones. In *Proceedings of the 20th International Workshop on Mobile Computing Systems and Applications (HotMobile '19)*. 135–140.

[22] X. Ran, C. Slocum, Y. Tsai, K. Apicharttrisorn, M. Gorlatova, and J. Chen. 2020. Multi-User Augmented Reality with Communication Efficient and Spatially Consistent Virtual Objects. In *Proceedings of the 16th International Conference on Emerging Networking EXperiments and Technologies (CoNEXT '20)*. 386–398.

[23] Remcom Wireless InSite 3D Wireless Prediction Software 2021. (2021). https://www.remcom.com/wireless-insite-em-propagation-software

[24] S. Sur, I. Pefkianakis, X. Zhang, and K. Kim. 2018. Towards Scalable and Ubiquitous Millimeter-Wave Wireless Networks. In *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking (MobiCom '18)*. 257–271.

[25] S. Sur, X. Zhang, P. Ramanathan, and R. Chandra. 2016. BeamSpy: Enabling Robust 60 GHz Links under Blockage. In *Proceedings of the 13th Usenix Conference on Networked Systems Design and Implementation (NSDI'16)*. 193–206.

[26] View Frustum Culling 2015. (2015). http://www.lighthouse3d.com/tutorials/view-frustumculling/

[27] S. Wang, J. Huang, X. Zhang, H. Kim, and S. Dey. 2020. X-Array: Approximating Omnidirectional Millimeter-Wave Coverage Using an Array of Phased Arrays. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking (MobiCom '20)*.

[28] Teng Wei and Xinyu Zhang. 2017. Pose Information Assisted 60 GHz Networks: Towards Seamless Coverage and Mobility Support. In *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking (MobiCom '17)*. 42–55.

[29] Teng Wei, Anfu Zhou, and Xinyu Zhang. 2017. Facilitating Robust 60 GHz Network Deployment by Sensing Ambient Reflectors. In *Proceedings of the 14th USENIX Conference on Networked Systems Design and Implementation (NSDI'17)*. 213–226.

[30] wil6210 module 2021. (2021). https://git.kernel.org/pub/scm/linux/kernel/git/kvalo/ath.git/tree/drivers/net/wireless/ath/wil6210

[31] T. Xu, B. Han, and F. Qian. 2019. Analyzing Viewport Prediction under Different VR Interactions. In *Proceedings of the 15th International Conference on Emerging Networking Experiments And Technologies (CoNEXT '19)*. 165–171.

[32] Z. Yang, P. H. Pathak, Y. Zeng, and P. Mohapatra. 2015. Sensor-Assisted Codebook-Based Beamforming for Mobility Management in 60 GHz WLANs. In *MASS '15*.

[33] X. Yin, A. Jindal, V. Sekar, and B. Sinopoli. 2015. A Control-Theoretic Approach for Dynamic Adaptive Video Streaming over HTTP. In *Proceedings of ACM SIGCOMM*.

[34] Anlan Zhang, Chendong Wang, Bo Han, and Feng Qian. 2021. Efficient Volumetric Video Streaming Through Super Resolution. In *Proceedings of the 22nd International Workshop on Mobile Computing Systems and Applications (HotMobile '21)*. 106–111.

[35] D. Zhang, M. Garude, and P. H. Pathak. 2018. MmChoir: Exploiting Joint Transmissions for Reliable 60GHz MmWave WLANs. In *Proceedings of the Eighteenth ACM International Symposium on Mobile Ad Hoc Networking and Computing (Mobihoc '18)*. 251–260.

[36] W. Zhang, B. Han, P. Hui, V. Gopalakrishnan, E. Zavesky, and F. Qian. 2018. CARS: Collaborative Augmented Reality for Socialization. In *Proceedings of International Workshop on Mobile Computing Systems and Applications (HotMobile)*.