# CMU-VP: Cooperative Multicast and Unicast With Viewport Prediction for VR Video Streaming in 5G H-CRAN

JUNCHAO YANG[1], JIANGTAO LUO[2], (Senior Member, IEEE), JUNXIA WANG[1], AND SONG GUO[3], (Senior Member, IEEE)

[1]School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
[2]Electronic Information and Networking Institute, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
[3]Department of Computing, The Hong Kong Polytechnic University, Hong Kong

Corresponding author: Jiangtao Luo (luojt@cqupt.edu.cn)

**ABSTRACT** Virtual reality (VR) is commonly regarded as one of 5G killer-applications. Transmission efficiency and quality of experience (QoE) are the most concerning issues for VR video streaming in 5G networks. Several multicast approaches were proposed to address these issues regardless of variance of personal viewports. In this paper, we explore a novel scheme combining multicast and unicast sessions in heterogeneous cloud-radio access networks (H-CRAN), in which a basic version of the video is transmitted to all users through the g-NB in a multicast session, and tiles of enhanced-version are transmitted to each viewer in a unicast session through its stationed remote radio head (RHH). To ensure the real-time content delivery, a user's viewport is predicted using a method based on historical trajectories and similarity of motion behavior, and then the tiles of predicted viewport in a version dependent on the channel quality are sent to the user in the unicast session. The scheme is formulated into a mixed-integer nonlinear problem (MINLP), and two near-optimal solutions are proposed to solve it by applying greedy approach and approximate approach, respectively. The simulation results show that our proposed scheme ensures better QoE under constrained bandwidth, and the proposed near-optimal solutions can efficiently solve the problem with low complexity and comparable performance.

**INDEX TERMS** VR video, quality of experience (QoE), resource allocation, cellular network, multicast, unicast, H-CRAN.

## I. INTRODUCTION

Recently, as the popularization of virtual reality (VR), increasing people are able to experience VR capabilities on affordable head-mounted display (HMDs) (e.g., HTC VIVE). VR makes use of 360-degree panoramic videos with high resolution (higher than 4K), high frame rate (60–90 fps) and low delay (less than 20 ms) to provide an immersive environment for the user to interact with the virtual world [1]. However, streaming bandwidth-intensive VR video over current network is quite challenging. First, compared to traditional video, its larger size and higher resolution lead to

intensive bandwidth demand. Second, unlike traditional video viewing, where screens are often shared (e.g., family watching TV together), HMD screens cannot be shared and each HMD requires the same content need to be delivered separately [2]. Third, compared to communication delay (source provider to user equipment) for current network (e.g., 10∼200ms for LTE), motion-to-photon (MTP) latency for VR is quite critical (i.e., less than 20ms) [2]. Therefore, the increase of end-to-end latency will degrade the QoE significantly [3]. Moreover, only portion of the panoramic VR video is watched at a certain time due to the limitation of viewer's field of view (FOV), which causes significant wasting of bandwidth resource. Consequently, streaming high quality 360-degree videos efficiently is one of the most critical issues for current VR applications.

The associate editor coordinating the review of this manuscript and approving it for publication was Xiaodong Xu.

Optimization of VR video transmission is actively under exploration. In order to reduce bandwidth demand, tile based adaptive streaming is proposed [4], which partitioned the panorama into independent tiles, enabling various content rate adaptable to the desired viewport of viewers. Based on tiling scheme, multiple approaches has been proposed to improve the quality of experience (QoE) [5], [6]. However, these approaches always transmit panorama videos to viewers in a unicast manner, e.g, streaming a VR video to millions of viewers, which is not efficient even though tiling can help with bandwidth reduction. Meanwhile, latency is another critical issue for adaptive streaming, since HMD updates the viewport data with a frequency between 60Hz and 120Hz [8] (which means the latency for data input should be less than 20ms). Hence, viewport prediction based prefetching strategy is generally adopted, which significantly reduce the latency as well as the transmission bandwidth. However, the accuracy of historical motion based prediction [10] will decrease sharply along the time. Once the prefetched quality-foveated tiles is mismatched with the real viewport, QoE of viewer will be degraded noticeably. Recently, a few attempts to utilize multicast scheme to serve a huge number of viewers consuming the same VR content simultaneously [7]–[9]. The multicast solutions significantly improve the efficiency of VR video transmission. However, it fails to fulfill individual demand of viewers with respective of viewport (i.e., one has to compromise to view low quality of VR video due to the user grouping for spectral efficiency improvement).

As for the underlying technology, 5G heterogeneous cloud-radio access networks (H-CRAN) has shown as a promising solution for VR video transmission in the near future [11]. At early stage of VR application, most of VR videos are downloaded in advance and viewed locally by moving the viewport in real-time. Recently, an increasing number of people prefer to experience VR at any time and any place through portable HMDs integrated with powerful Graphics Processing Unit (GPU), which is also referred as Mobile VR. Unsurprisingly, Mobile VR is predicted as one of 5G killer-applications [3]. H-CRAN facilitated with mobile-edge computing (MEC) [12], together with dual radio architecture [22] are envisioned to be typical in 5G systems. Heterogeneous networks with massive densification of small cells and C-RANs are combined in one network structure to improve spectral efficiency, resource management and energy efficiency. Naturally, those make bandwidth-intensive VR video streaming over H-CRAN feasible.

In 5G H-CRAN, dual connectivity (DC) allows users to be simultaneously served by a macro and a small cell operating at different carrier frequencies [10]. By exploiting the DC, a multimedia multicast transmission scheme through macro/small BS cooperation is presented in [10], in which multicast streaming and patching streaming is delivered by macrocell base station (MBS) and small-cell BS (SBS), respectively. Notice that the multicast can significantly improve the transmission efficiency but it will degrade the viewers' QoE. By contrast, the QoE can be guaranteed

by unicast but it will result in ineffective transmission, especially when considerable number of people are request same VR content simultaneously (e.g. a popular VR film). Is it possible to transmit VR video by taking advantage of multicast for transmission efficiency and unicast for QoE improvement ? Motivated by this idea, in this paper, we explore VR video streaming in 5G H-CRAN through cooperative multicast and unicast (CMU) by macro (i.e., g-NB) and small cell (i.e., RRH), respectively. More specifically, a certain version of panoramic tiled VR video is transmitted to all viewers through g-NB in a multicast manner. Meanwhile, based on cross user viewport prediction, a portion of enhanced tiles in the predicted viewport is transmitted to respective viewers in a unicast manner through corresponding RRH. In order to improve the accuracy of viewport prediction for unicasting enhanced tiles, we explore the viewport prediction based on historical trajectories and similarity of cross viewer motion behavior. Furthermore, with proactive caching and real-time transcoding paradigms in MEC, the end-to-end latency is reduced significantly as that pushes the data close to the user equipments (UEs).

Our main contributions are described as follows:

- In order to achieve better transmission efficiency and QoE improvement, we proposed the cooperative multicast and unicast with viewport prediction (CMU-VP) scheme for VR video streaming in H-CRAN.
- Viewport prediction based on historical trajectories and similarity of cross viewer motion behavior is proposed for better accuracy.
- To solve the mixed-integer nonlinear problem (MINLP), two near-optimal solutions with low complexity and comparable performance are presented by exploring the greedy approach and approximate approach, respectively.
- Plenty simulations are conducted to verify that CMU-VP scheme ensures better QoE under limited bandwidth, and the near-optimal solutions can efficiently solve MINLP with low complexity and comparable performance.

The remainder of the paper is organized as follows. Section II provides an overview of related work. System model is presented in Section III. Cross user viewport prediction and proposed CMU-VP scheme are described in Section IV. Problem formulation and Solutions are represented in Section V. Evaluations are performed in Section V. Finally, Section VI concludes the paper.

## II. RELATED WORK
### A. ADAPTIVE VR STREAMING
Aiming at reducing the bandwidth consumption, video coding solutions are widely adopted for VR video streaming in the literature. Among these approaches, tile-based and viewport based are the most common approaches for adaptive streaming, which common strategy is to transmit portion of panoramic video in high quality while the

rest in lower quality. Tile based adaptive streaming is first proposed by Patrice *et al.* [4], in which a greedy approach is adopted to solve the optimization of tile selection. Furthermore, the impact of tiling scheme on quality of transmission is evaluated. Based on tiling scheme, multiple approaches has been proposed to improve QoE [5], [6].

Viewport based adaptive streaming is typical represented by Facebook, a VR video is transformed and encoded into multiple versions towards different perspectives. And a viewer requests one of the video versions according to the orientation. However, such scheme produces high redundancy of contents (e.g., Facebook creates 150 versions for one VR video) which require huge amount of storage and bandwidth.

### B. MOBILE VR STREAMING

The previous studies focused on wireline VR video transmission or offline delivery for local viewing on HMD. However, with the growing capacity of mobile network and increasingly popular of Mobile VR, the demand of transmission high quality VR video efficiently over mobile network is becoming ever urgent [15]. To cater for data intensive VR video delivery over mobile network, Qian *et al.* [16] proposed a cellular-friendly VR video streaming scheme that transmits only visible portion of video based on head movement prediction. Similar approach is also presented for mobile VR [14], which improvement is to complement the viewport based method by adding foveated provision within each view. While these approaches heavily depend on viewport prediction, the QoE will degrade dramatically once the transmitted data mismatch the real Field of View (FOV). Furthermore, the chunks of data stream has to be segmented relatively short to keep peace with frequent viewer's viewport changes. Consequently, the efficiency of transmission with short chunks is low because of ineffective compression and heavy cost of data synchronization. An efficient User-Generated Content (UGC) VR video transmission scheme over cellular network is represented in our previous work [25], in which only one representation of each tile is generated for uploading based on optimization of uplink resource allocation under the consideration of quality of content (QoC) contribution, then directly transmitted to viewer without transcoding.

However, transmitting a bandwidth-intensive VR video to multiple users over resource-limited mobile network in unicast manner is an inefficient way. Hence, a few researches explore VR video multicast for the objective of transmission efficiency. Bao *et al.* [8] designed a bandwidth-efficient scheme that combines multicast and unicast by sending partial of VR video based on motion prediction, to decide whether to use one multicast or multiple unicasts when a certain viewing area on the sphere is required by multiple viewers. In addition, margins are added to the predicted FOV in order to handle prediction errors. In order to improve spectral efficiency of VR video multicast. Jounsup *et al.* [7] proposed a multi-session multicast approach for VR video, in which user grouping, wireless resource allocation and tile

**TABLE 1.** List of symbols.

| Symbol | Quantity |
|--------|----------|
| $T$ | Number of tiles |
| $t$ | Index of tile |
| $V$ | Number of viewer in the system |
| $v$ | Index of viewer |
| $n$ | Index of RRH |
| $N$ | Number of RRHs |
| $Q$ | Number of tile representations |
| $K$ | Number of users with similar motion |
| $R$ | Data rate of Tile |
| $U$ | Utility function |
| $D$ | Time window of similarity measurement |
| $\Omega$ | Overlapped tiles between two users |
| $\Theta$ | Number of tile in viewport |
| $\alpha, \beta$ | Coefficients of the utility model |
| $x_{v,t}$ | The indicator for t th tile unicast to v th viewer or not |
| $p_{v,t}$ | The viewing probability of t th tile by v th viewer |
| $O_{LR}$ | The estimated fixation predicted by linear regression |
| $O_{SM}^{k}$ | The fixation of K most similar cross-users |
| $\gamma_{LR}$ | The regression coefficient of linear regression model |
| $W_{LR}$ | The weight of LR prediction |
| $W_{SM}$ | The weight of similar user fixation |
| $\xi_t$ | The votes for a certain tile |

rate selection were jointly optimized with spectral efficiency maximization. Inspired by the similar motivation, a multicast DASH-based tiled streaming solution was presented in [2], and a heuristic algorithm was proposed to solve the tile rate adaptation problem. In addition, a QoE-aware deep learning-aided VR multicast framework was presented by Cristina Perfecto *et al.* [17], in which the future FoV of VR users were predicted by a deep neural network. Based on prediction FOV, proactive multicast resource scheduling was performed. And the problem was formulated as a request admission maximization problem and was solved by a low complexity matching algorithm. Furthermore, Athul Prasad *et al.* [18] discussed the challenges for VR broadcast using 5G small cell network, solutions in terms of usage of single frequency network (SFN) type of deployments and unlicensed millimeter wave (mmW) bands were considered. Recently, they proposed a D2D assisted VR broadcast scheme that enabled radio resource efficient delivery VR video using broadcast transmission [19].

## III. SYSTEM MODEL

The system considered in this work is depicted in Fig. 1, which consists of three part: VR content Server, MEC Server and H-CRAN. The detail of the system are described as follows:

The VR content Server is responsible for storing the encoded VR video. With motion-constrained tile sets (MCTS) coding [5], the VR video is spatially split into rectangular, independently decodable, non-overlapping tiles after Equirectangular Projection (ERP) [8] to facilitate viewport cropping, which are denoted by $t \in T$. And the stream of each tile is then segmented into chunks temporally. Generally, multiple representations for each tile are generated at the content Server to facilitate further adaptive streaming, and the
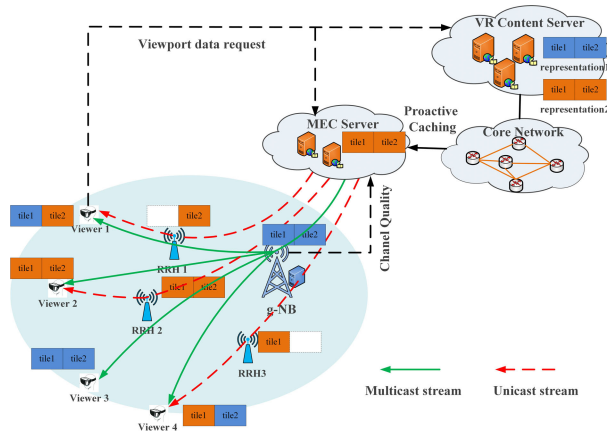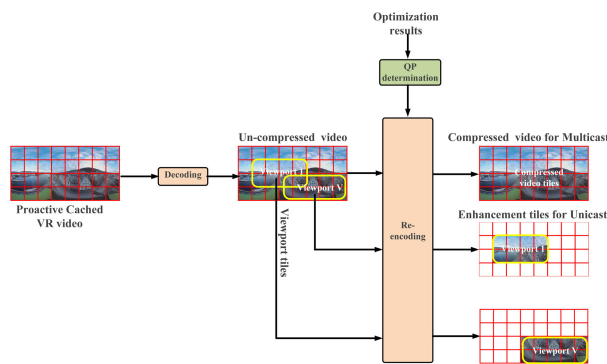
**FIGURE 1.** VR video streaming in 5G H-CRAN.



**FIGURE 2.** MEC-based transcoding module for CMU.

are connected to a centralized baseband-unit (BBU) pool through backhaul links and fronthaul links, respectively [11]. The BBU pool executes upper layer functions and baseband signal-processing, whereas, the RRHs perform as radio-frequency (RF) transceivers and only perform basic RF functions. Software defined network (SDN) is adopted to support the separation of data and control information, which divides the control information into g-NB and data information into RRHs [12]. The g-NB is mainly responsible for the delivery of control information and multicast service. For the simplicity of index, the RRHs and g-NB are uniformly denoted by $n \in N$. To protect the control information and multicast signal, spectrum of g-NB and RRHs are assigned at different frequencies. In addition, the spectrum assigned to RRHs are reused at each RRH. Moreover, cyclic prefix OFDM (CP-OFDM) [11] is considered as multiple access technique in our presented architecture.

Furthermore, consider $V$ randomly viewers consuming one VR video simultaneously. DC is supported for all user equipments (UEs) (i.e., viewer HMDs in our system). The signal interference to noise ratio (SINR) corresponding to unicast link and multicast link are estimated from the uplink sounding reference signals that are periodically broadcasted from the UE. Based on the link quality and predicted viewport information, optimization of CMU scheduling is performed at the MEC server. The optimization results will be sent to the UE, and the UE request the appropriate quality of viewport tiles and entire transcoded VR video according to the optimization result. Finally, the MEC server performs the transcoding task and send the proper chunk of tiles by CMU through g-NB and RRH, respectively.

## IV. CROSS USER VIEWPORT PREDICTION AND PROPOSED CMU-VP SCHEME
### A. CROSS USER VIEWPORT PREDICTION
Cross user viewport prediction is proposed, in which, K most similar users is first selected based on similarity of historical motion, then the K users' fixation is used to amend the prediction result achieved by linear regression approach.

Tile based adaptive streaming has shown as a promising way to transmit VR video efficiently [4]. However, due to the delay sensitive feature of VR application (e.g., motion to photon latency should be less than 20 ms), it is highly impracticable to transmit high quality of video at current viewport area immediately. Thus, viewers has to keep relative large buffer for proactively pre-fetching tiles to ensure continuous playback, and it is necessary to predict the viewport for the reason of quality enhancement on portion of pre-fetching tiles. Motion-based [8] and content-based [13] approaches are most widely adopted for viewport prediction. However, prediction accuracy of motion-based approach drops quickly along with the time, and content-based approach is probably not reliable since how texture and motion of video content influences on the user motion are not fully investigated [20]. Recently, Ban *et al.* [14] exploited cross user behavior to predict viewport by K-Nearest-Neighbors (KNN) method,

bitrate representations are denoted as $\{R_t^1, ..R_t^Q\}$, where $t$ is represented the tile index, $R_t^1$ and $R_t^Q$ represent the predefined minimum and maximum encoding rate of $t$-th tile, respectively. In addition, a cross user viewport prediction module is designed at the VR content sever to utilize the similarity of user viewport motion, then the result of prediction is fed back to MEC server for CMU scheduling.

The MEC server is located at the edge of H-CRAN, which enables proactive caching and real-time transcoding for request content. When users within the coverage of H-CRAN request the VR content from the content server, the request data will be retrieved from the MEC server. We assume that the only one representation of a VR video with the highest quality is proactively cached at the MEC server. To provide appropriate tile representations for CMU, a real-time transcoding module is designed in this paper, as shown in Fig. 2. Viewport tiles for each unicast links and transcoded entire VR video for multicast link are re-encoded based on the proactive cached representation, and the appropriate quantization parameters (QPs) for unicast stream and multicast stream are selected according to the optimization of CMU scheduling.

Two-tier H-CRAN are considered in this work, where a macro 5G base station (i.e., g-NB) are underlaid with small cells (i.e., RRHs), the g-NB and small cell RRHs
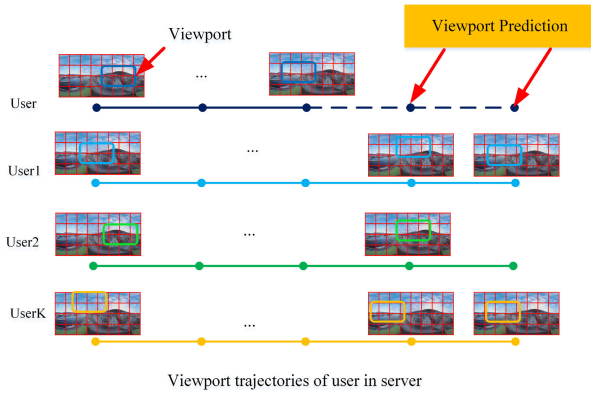
**FIGURE 3.** Cross user viewport prediction.

which achieved considerable performance. However, KNN method fails to consider the similarity between cross user motion behaviors, since the K nearest fixation of other users is used to amend the prediction result based on historical trajectories. While motion behavior of the prediction user is probably quite different from the K users' (i.e., each user can potentially view a VR video in a unique fashion, under this situation, using the K nearest fixation to amend the prediction result will lead to larger errors). On the other hand, some research have already shown that users in virtual environment have certain similar viewing patterns when watching the same VR video [21].

Based on the aforementioned analysis, we intend to utilize the similarity of user motion behaviors for higher accuracy of viewport prediction. The demonstration of proposed prediction approach is shown as Fig.3, where the viewport is first predicated based on historical trajectories, then the prediction result is amended by exploiting similarity of motion behaviors of cross users. The similarity of motion behaviors between two user is defined as (1), where $t_0$ represents system time, $\delta$ is sampling index of historical trajectories, $D$ is the time window of similarity measurement, $\Omega_{t_\delta}\{v, j\}$ represents the overlapped number tiles between two users' viewport at time $t_\delta$. $v$ and $j$ are the index of two different user, and $v, j \in V$. $\Theta_{t_\delta}\{v\}$ represents number of tiles in viewer's viewport at time $t_\delta$.

$$Sim\{v, j\} = \sum_{\delta=0}^{D} \frac{\Omega_{t_\delta}\{v, j\}}{\Theta_{t_\delta}\{v\} + \Theta_{t_\delta}\{j\}} \quad (1)$$

Similar strategy as KNN approach is adopted in our work, top K of most similar motion behaviors of users are selected for amending the historical trajectories based prediction.

For historical trajectories based prediction, a linear regression (LR) model in window $(S_0 - 1, S_0)$ is trained to predict viewport in future time $t_0 + \Gamma$, which can be formulated as (2), where $S_0$ and $\Gamma$ denote the chunk index at $t_0$ and the time duration of chunk, respectively. $O_{LR}$ represents the estimated fixation predicted by LR, and $\omega_{LR}$ is the regression coefficient of LR.

$$O_{LR}(t_0+\Gamma) = \omega_{LR}\Gamma + O_{LR}(t_0) \quad (2)$$

We adopt the vote mechanism which considering fixation weights and tile's viewed times by the K of most similar users. Furthermore, the votes for a certain tile $\xi_t$ can be written as (3), where $W_{LR}$ and $W_{SM}$ represent weights of LR prediction and weights of similar user fixation, respectively. $O_{SM}^k$ denotes the fixations of K most similar cross-users. $F(O)$ indicates the covering field of each fixation represented by a $T$-dimension vector, where $T$ represents the tile number indexed in raster-scan order, and $F(O) = 1$ means the tile is viewed, 0 otherwise.

$$\xi_t = W_{LR} \cdot F_t(O_{LR}) + \sum_{k}^{K} W_{SM} \cdot F_t(O_{SM}^k) \quad (3)$$

Finally, viewing probability of each tile $p_t$ for the user can be written as (4), i.e.,normalizing votes for each tile.

$$p_t = \frac{\xi_t}{\sum_{t}^{T} \xi_t} \quad (4)$$

### B. PROPOSED CMU-VP SCHEME

The scheme with reasonable tradeoff between transmission efficiency and QoE is proposed for efficient VR video streaming. Streaming entire tiled video to users in unicast manner wastes significant network bandwidth, because users can only watch a portion of video at a time through the HMDs, while the rest of transmitted video outside the user viewport is wasted. Multicasting VR content to large number of users who consuming same VR video simultaneously is an efficient way. However, it would lead to part of users with lower QoE, since it fails to enhance content quality within respective viewport of users. To utilize DC feature in H-CRAN, we take both advantages of unicast on QoE improvement and multicast on transmission efficiency, i.e. cooperative multicast and unicast with viewport prediction, which is referred as CMU-VP in this paper. Specifically, a certain version of entire tiled VR content is transmitted to all viewers through g-NB in a multicast manner, while portion of enhancement tiles are transmitted to each viewer based on predicted viewport through corresponding RRH in a unicast manner. For each viewer, after receiving multicast and unicast data, then merges them into a new version of VR content for future display.

The detail of CMU-VP is summarized as follows: 1)Tiled VR video with the highest representation is proactively cached at the MEC server, when users within the coverage of H-CRAN request the content from the content server, the request data will be retrieved from the MEC server. 2) Meanwhile, viewport prediction is performed at the VR content sever according to our proposed cross user viewport prediction method, then the result of prediction is fed back to MEC server for CMU scheduling. 3) Once VR content server receive the consuming requests of VR video from viewers, it creates a multicast service at Broadcast Multicast Service Center (BM-SC), which is responsible for multicast sessions management [23]. Meanwhile, a unicast stream

is also maintained with each UE for transmitting enhancement tiles. 4) Optimization of CMU scheduling are performed jointly considering probability of tile viewing, the channel quality of multicast link and unicast links. Then a certain version of entire tiled video and enhancement tiles responding to each predicted viewport are generated through the real-time transcoding. 5) With the assistance of SDN, the certain version of entire tiled video and enhancement tiles are distributed into g-NB for multicast and RRHs for unicast, respectively. 6) With DC supported, UEs receive multicast stream from g-NB and unicast stream from corresponding RRH. Finally, multicast and unicast streams are merged (i.e., the lower representation of the tiles received from both unicast and multicast streams is discarded) in buffer of UEs to achieve a high quality of VR content.

Note that the objective of CMU-VP is to optimize QoE for all the viewers. Thus, the version of entire tiled video (i.e., tiles with which representations) for multicast, and unicast enhancement tiles (i.e., which tiles and with which representations) to respective viewer need to be determined for that purpose. Thus, the problem of VR video streaming through CMU-VP is formulated as a QoE optimization problem, and detail of problem formulation and solution are given in next section.

## V. PROBLEM FORMULATION AND SOLUTIONS
### A. PROBLEM FORMULATION

In this paper, we adopt QoE metric for VR video streaming in [7], and utility of a tile is defined as (5), where $R_{v,t}$ and $R_t^Q$ represent the encoding rate of $t$-th tile transmitted to $v$-th UE and the pre-defined maximum tile encoding rate, respectively. And $\alpha$ and $\beta$ are the coefficients of utility model. The utility is strictly concave function of received video rate, and also marginally decreasing as the increase of video rate. These features of utility function which can model the quality of user experience well as the received video rate increasing. For tile based VR video, the utility of a tile is used to model the contribution of tile with certain bitrate to the whole quality of user experience.

$$U_{v,t} = \begin{cases} \alpha \log(\beta R_{v,t}/R_t^Q), & R_t^{v,t} > 0 \\ 0, & R_{v,t} = 0 \end{cases} \quad (5)$$

According to conventional multicast scheme [7], the transmission rate of a group is determined by the user with the worst channel condition in that group. Thus, the multicast rate can be written as follows:

$$r^m = B^m log_2 \left(1 + \min_{v \in V}(\frac{P_0 E\left\{|h_{0,v}|^2\right\}}{\sigma^2})\right) \quad (6)$$

where $B^m$ represents bandwidth for multicast, $P_0$ represents transmission power of g-NB, $E\left\{|h_{0,v}|^2\right\}$ is the UE $v$ average channel gain from g-NB, and $\sigma^2$ is the power spectral density of additive white Gaussian noise. Note that, in order to ensure

the transmission efficiency, only one user group is considered in our scheme.

And the unicast rate of UE $v$ associated $n$ RRH can be written as follows:

$$r_v^u = B^n \log_2 \left(1 + \frac{P_n E\left\{|h_{n,v}|^2\right\}}{\sigma^2}\right) \quad (7)$$

where $B^n$ represents available bandwidth of RHH $n$, $P_n$ represents transmission power of RHH $n$, $E\left\{|h_{n,v}|^2\right\}$ is the UE $v$ average channel gain from RHH $n$.

Let $x_{v,t} \in \{0, 1\}$ indicates whether tile $t$ is transmitted to viewer $v$ through unicasting. $R_{v,t}^u$ denotes the representation of tile $t$ transmitted to viewer $v$ through unicasting, and $R_{v,t}^m$ represents multicast representation of tile $t$. Note that each viewer always chooses a better representation of tile if the tile is received by viewer from both unicast and multicast streams. We adopt weighted utility function [4] to model the expected QoE for VR video streaming in this paper, the utility of a tile is used to model the contribution of tile with certain bitrate to the whole quality of user experience, and weight $p_{v,t}$ is the probability of tile $t$ visible in the viewport of viewer $v$. In order to maximize the whole QoE of all users under limited bandwidth, the version of entire tiled video for multicast (i.e., $R_{v,t}^m$), and unicast which enhancement tiles with which bitrate (i.e., $x_{v,t}$, $R_{v,t}^u$) to respective viewer need to be determined by CMU-VP scheduling. Consequently, CMU-VP schedule problem with the objective of QoE maximization for all viewers can be formulated as below:

$$P1: \max_{x_{v,t},R_{v,t}^u,R_{v,t}^m} \sum_v^V \sum_t^T \left(x_{v,t}U_{v,t}^u+(1-x_{v,t})U_{v,t}^m\right)p_{v,t}$$

$$s.t. \sum_t^T R_{v,t}^m \leq r^m, \quad \forall v \in V, \forall t \in T \quad (8)$$

$$\sum_t^T R_{v,t}^u \leq r_v^u, \quad \forall v \in V, \forall t \in T \quad (9)$$

$$R_{v,t}^u, R_{v,t}^m \in \{R_t^1, ..R_t^Q\}, \quad \forall v \in V, \forall t \in T \quad (10)$$

$$x_{v,t} \in \{0, 1\}, \quad \forall v \in V, \forall t \in T \quad (11)$$

The constraint (8) limits the sum of multicast tiles rate in bits/sec is less than the rate achieved from g-NB for multicasting. (9) indicates that total transmitted tiles rate to each viewer in a unicast manner should be less than the achievable rate of each UE from associated RRH. (10) limits the muticast and unicast tile rate is selected from pre-defined tile representations, respectively. (11) is binary constraint indicates whether tile $t$ is transmitted to viewer $v$ in a unicast manner.

### B. SOLUTIONS

The problem $P1$ is a MINLP, and is proved NP-hard in [24]. Due to high computation complexity of exhaustive search and heuristics algorithms, we explore to derive two near-optimal solutions for the problem, and details are given as follows.

---

**Algorithm 1** Greedy Algorithm

---

1: $\Theta_v$: Predicted viewport of $v$ th viewer
2: **for** $v = 1; v \leq V; v + +$ **do**
3:  **for** $t = 1; t \leq T; t + +$ **do**
4:   // determine $x_{v,t}$ according to predicted viewport
5:   **if** $t \in \Theta_v$ **then**
6:    $x_{v,t} = 1$
7:   **else**
8:    $x_{v,t} = 0$
9:   **end if**
10:  **end for**
11: **end for**
12: **for** $v = 1; v \leq V; v + +$ **do**
13:  **if** $x_{v,t} = 1$ **then**
14:   // determine $R_{v,t}^u$
15:   apply greedy approach to solve $P2$
16:  **else**
17:   // determine $R_{v,t}^m$
18:   apply greedy approach to solve $P3$
19:  **end if**
20: **end for**

---

### 1) GREEDY ALGORITHM

Greedy algorithm solve the problem by adopting greedy approach after one variable $x_{v,t}$ is firstly determined according to the viewport prediction. Note that there are three variables(i.e., $x_{v,t}$, $R_{v,t}^u$, $R_{v,t}^m$) need to be determined in the $P1$, it's highly impossible to simultaneously find the optimal solutions for objective optimization. However, we notice that once the $x_{v,t}$ is determined, and $R_{v,t}^u$, $R_{v,t}^m$ can be selected by solving the $P2$ and $P3$, respectively. Fortunately, $x_{v,t}$ for each viewer can be first determined based on viewport prediction as we assuming that the FOV consists constant number of tiles for each viewer. With the highest viewing probability tile index, the tiles within the the predicated viewport is selected for unicast to viewer.

With determined $x_{v,t}$, for each viewer, the unicast and multicast schedule problem can be derived into two independent sub-problems $P2$ and $P3$. While the similar problem has been studied in our previous work [25], in which we adopted the greedy approach to solve this type of problem. Therefore, the problem $P1$ is finally solved by applying the greedy approach, and the Algorithm 1 summarizes the details of our approach, which is given as following:

$$P2 : \underset{R_{v,t}^u}{Max} \sum_v^V \sum_t^T x_{v,t} U_{v,t}^u p_{v,t}$$
$$s.t.(9), (10)$$
$$P3 : \underset{R_{v,t}^m}{Max} \sum_v^V \sum_t^T (1 - x_{v,t}) U_{v,t}^m p_{v,t}$$
$$s.t.(8), (10)$$

It should be noted that Algorithm 1 achieves sub-optimal solutions as we roughly derive $P1$ into two independent

problems $P2$ and $P3$ after the $x_{v,t}$ is determined. In addition, utilizing the greedy approach to solve $P2$ and $P3$ separately will also lead local optimal. Meanwhile, the computational complexity of greedy approach is relatively high.

### 2) DECOMPOSING ALGORITHM

Decomposing algorithm solves the problem by decomposing the $P1$ into two sub-problems. Note that, multicast and unicast jointly scheduling makes the problem $P1$ prohibitively difficult to solve. However, if the multicast schedule result is first obtained, then the unicast schedule problem with the multicast schedule result would be relatively easy to solve. Inspired by this idea, we decompose the problem $P1$ into two sub-problems $P4$ and $P5$, which can be described as multicast schedule problem and unicast schedule problem with the multicast schedule result, respectively.

$$P4 : \underset{R_t^m}{Max} \sum_t^T U_t^m \overline{p_t}$$
$$s.t. \sum_t^T R_t^m \leq r^m, \quad \forall t \in T \quad (12)$$
$$R_t^m \in \{R_t^1, ..R_t^Q\}, \quad \forall t \in T \quad (13)$$

where $\overline{p_t}$ is the average probability of tile $t$ for all multicast viewers, and $R_t^m = R_{v,t}^m$ indicates that all the viewers received same tile representation through the multicast stream.

$$P5 : \underset{x_{v,t}, R_{v,t}^u}{Max} \sum_v^V \sum_t^T \left( x_{v,t} U_{v,t}^u + (1 - x_{v,t}) U_{v,t}^m \right) p_{v,t}$$
$$s.t. (9), (10), (11)$$

The sub-problem $P4$ is a convex problem, which can be solved efficiently by optimization toolbox or greedy algorithm in our previous study [25].

After solving $P4$, $P5$ becomes a resource constrained knapsack problem. The $P5$ can be solved by relaxing it into a fractional knapsack problem then applying the greedy approach [4]. Due to the computationally fast and elitism strategy, we adopt APMonitor Modeling Language to solve $P5$ [26], which provides real-time optimization, and is freely available through MATLAB or Python interface. And APOPT solver provided by APM solves the problem $P5$ with obtained solution of $P4$ efficiently.

Note that more sophisticated algorithms can be applied to solve the $P1$ optimally, however, it should be high computation complexity cost. In contrast, according to our strategy of problem solutions, the complexity is greatly reduced and is easy to implement by adopting exist algorithm or optimization toolbox. Furthermore, simulations also show that the gap between proposed near-optimal solutions and optimal solution is relatively small, and near-optimal solutions can solve the problem efficiently with comparable performance. And details will be explained in next section.

**TABLE 2.** Parameters of simulation.

| Parameters | Value |
|---|---|
| g-NB number | 1 |
| RRH number | 3 |
| g-NB radius | 500m |
| RRH radius | 100m |
| g-NB Tx power | 46dBm |
| RRH Tx power | 30dBm |
| g-NB Tx bandwidth | [6,9,12] MHz |
| RRH Tx bandwidth | [4,6,8] MHz |
| Noise spectral density | -173 dBm/Hz |
| Shadowing standard deviation | 8 dBm |
| Path loss for g-NB | 140.7+36.7*log10(d) |
| Path loss for RRH | 128.1+37.6*log10(d) |



**FIGURE 4.** Comparison of viewport prediction accuracy.

## VI. SIMULATIONS

In this section, we conduct plenty simulations to evaluate the performance of our proposed solutions for the problem.

### A. SETUP

A g-NB cell with 0.5 km radius is considered in the simulation, where three RRHs is uniformly distributed within g-NB cell. The users' locations are randomly generated and UE follow a Poisson Point Process (PPP) with density distributed under g-NB coverage. The transmit power of g-NB and RHHs are 46 dBm and 30 dBm, respectively, and system bandwidth is set 10 MHz, 15 MHz, 20 MHz in different simulation conditions,respectively. Note that bandwidth is divided into two parts. one part is for multicast through g-NB, and the rest part is for unicast through RHHs, respectively (i.e., 6,9,12 MHz and 4,6,8 MHz in our simulation). And 60% of g-NB bandwidth is used for multicast service [7]. Lognormal shadowing with 8 dB standard deviation is implemented. The noise power spectral density is assumed to be -173 dBm/Hz, and more detail of simulation parameters is given in Table.2.

In order to fairly evaluate the proposed CMU-VP scheme with near-optimal solutions (**CMU Greedy**, **CMU Decomposing**), few existing schemes are adopted for comparison: 1) CMU optimal solution with exhaustive search (**CMU optimal**): CMU-VP scheme is adopted, and solved by exhaustive search to obtain optimal solution. 2) CMU-VP with equal tile bit rate (**CMU equal**): CMU-VP scheme is adopted, while tile bit rate is distributed equally without considering probability of viewing. 3) **Multicast only** [2]: Tile based multicast scheme according to [2], which jointly considers bandwidth constraint and tile weight. 4) **Unicast only** [5]: Tile based unicast scheme according to [5], where tile selection is based on viewport prediction.

4K resolution VR video (i.e., Freestyle Skiing) in [27] with 32 tiles is used for the simulation, and each tile is encoded by open-source HEVC encoder Kvazaar with 10 different representations: {0.1, 0.3, 0.5, 0.7, 0.9, 1, 1.2, 1.5, 1.7, 2.0} Mbps. And viewport trajectories of 48 users in dataset [27] are divided into two part, the first part are used for viewport prediction, while the other are adopted for performance evaluation.
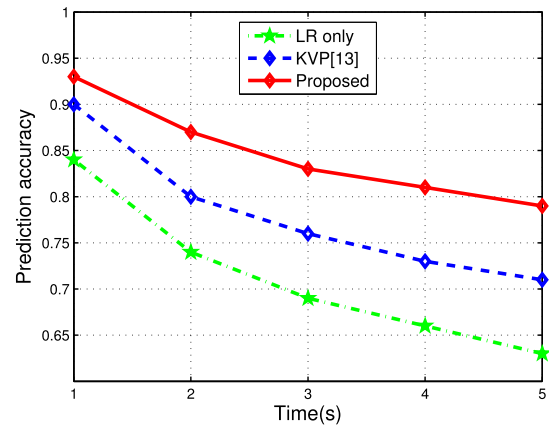
For one certain user, K most similar users is selected through the similarity measurement based on the first part of viewport trajectories. meanwhile, viewport prediction for next chunk is performed by LR based on historical trajectory. viewport prediction is performed by our cross user viewport prediction method. Then different streaming schemes are performed with viewing probability of each tile $p_t$ for each user. Finally, based on the second part of viewport trajectories, we evaluates the QoE objectively according to tile bitrate within each user's viewport.

### B. RESULTS

First, proposed viewport prediction method based on historical trajectories and similarity between cross user motion behaviors is evaluated. And conventional prediction method linear regression (LR) and KNN-based Viewport Prediction algorithm (KVP) in [14] are considered for comparison. For fairness of comparison, the value of K in our proposed method is set as the same in [14], i.e., $K = 5$. Fig.4 shows that accuracy of viewport prediction with different method along the time. LR method performs the worst, especially for long-term prediction. Because it performs prediction only based on historical data, and viewer motion behavior probably become quite different from that in former moment due to changes of content. Compared to KVP, our proposed method improves at least 6% higher accuracy, and maintains high accuracy over 80% even for the long-term prediction.

Fig.5 shows that normalized utility with different schemes in different bandwidth conditions. We can observe that the CMU-VP scheme always perform better than that without CMU-VP scheme (e.g., Multicast only and Unicast only approaches) in all bandwidth conditions. The Multicast only approach performs worst among all the approaches, because low tile bit rate is transmitted to viewers with the limited multicast bandwidth. The utility of Multicast only approach in average is 25% less than that achieved by CMU-VP optimal. Compared to CMU-VP optimal approach, CMU-VP proposed approach achieved comparable performance. While CMU-VP equal performed much worse, about 13% less
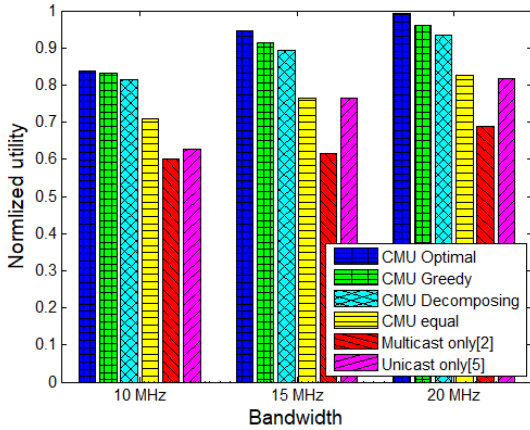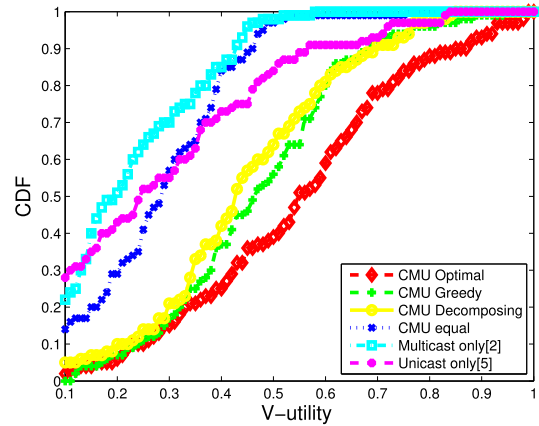
**FIGURE 5.** Normalized utility with $V = 5$.



**FIGURE 6.** Total UEs data rate over bandwidth with $B = 10$.



**FIGURE 7.** CDF of V-utility with $B = 10$, $V = 5$.

**TABLE 3.** Comparisons of complexities.

| Algorithm | Complexity | CPU time(sec) |
|---|---|---|
| Greedy | $O(VT + Q^{VT} + Q^T)$ | 0.42 |
| Decomposing | $O(Q^T + 2^{VTQ})$ | 0.23 |
| Exhaustive search | $O(Q^T \times 2^{VTQ})$ | 2.07 |

Then, cumulative distribution function (CDF) of tile utility in corresponding viewers' viewport (V-utility) is measured to evaluate the QoE objectively. As can be seen from Fig.7, approaches with CMU perform better than these without CMU. Furthermore, The gap between CMU optimal and CMU proposed indicates that the proposed near-optimal solutions are promising solutions with low complexity. Whereas, due to lack of transmission bandwidth, we can see that 30% of V-utility with Multicast only approach is below 0.1.

Finally, complexities of proposed algorithms are analyzed, which are presented in Table.3. Complexity of the greedy algorithm consists three part, the first part is the complexity for $x_{v,t}$ determination, and the second and third part are the complexity of greedy approach for $P2$ and $P3$, respectively. Complexity of Decomposing algorithm is sum of the complexity of greedy approach for $P4$ and $P5$ in the worst case, and the complexity of approximate approach(i.e., AMP) for $P5$ should be less than that with greedy approach. Exhaustive search solve the problem by searching all the possible options, therefore, the complexity is in order of multicast tile rate select options times unicast enhancement tile options.

Table.3 shows the comparisons of complexities. CPU time is also measured using Matlab in the condition with 10 MHz system bandwidth. The greedy algorithm and Decomposing algorithm take shorter time to complete the optimization than the exhaustive search approach. The Decomposing algorithm algorithm takes shortest time to obtain the solution, however achieved utility is less than greedy algorithm.

Overall, the proposed scheme could significantly improve QoE by utilizing the heterogeneous feature of H-CRAN. And the proposed near-optimal algorithms (i.e., greedy algorithm and decomposing algorithm) are promising solutions with low complexity and comparable performance.

utility compared with CMU-VP optimal. As the resources for transmission are equally distributed to each tile without considering probability of viewing, which leads the tile within the viewport in low quality (i.e., low V-utility value). And Unicast only performs steadily as the bandwidth increasing, however, much of transmission bandwidth is wasted to transmit entire of VR video to respective viewers in a unicast manner, which leads to the worst efficiency.

Fig.6 shows the total achieved data rate of UEs over total transmission bandwidth, which reflects the efficiency of transmission schemes. Clearly, We can observe that schemes with CMU(i.e., CMU optimal, CMU Greedy, CMU Decomposing, CMU equal) outperform schemes without CMU, since the bandwidth resources are fully utilized for schemes with CMU. And the unicast only performs the worst. Note that the only difference among the schemes with CMU is the strategy of tile rate allocation. However, tile rate allocation is based on the achieved data rate of UEs, which is same for schemes with CMU. Therefore, schemes with CMU has same performance on transmission efficiency. In fact, the curves of schemes with CMU are overlapped (i.e., CMU optimal in red, CMU Greedy in green, CMU Decomposing in yellow, CMU equal in blue) in Fig. 6.

## VII. CONCLUSION

In this paper, we explore viewport prediction based on historical trajectories and similarity between cross user motion behaviors. In order to achieve better transmission efficiency and QoE improvement, we propose the CMU-VP scheme for VR video streaming in H-CRAN. Furthermore, two near-optimal solutions for CMU-VP schedule problem is presented. The simulation shows that the proposed CMU-VP scheme with our proposed viewport prediction method can improve at least 25% and 17% higher QoE compared with muticast only scheme and unicast only scheme, respectively. Notice that time window of similarity measurement in our prediction method is a very important factor, which affects the accuracy of prediction. It's worth to careful study for long-term viewport prediction. CMU-VP might be a promising way to stream VR video to a group of people in some specific scenarios. For CMU scheduling, more sophisticated algorithm with low complexity will be explored in our future work.

## REFERENCES

[1] C. Westphal, "Challenges in networking to support augmented reality and virtual reality," in *Proc. IEEE Int. Conf. Comput., Netw. Commun. (ICNC)*, Sacramento, CA, USA, 2017, pp. 1–5.

[2] H. Ahmadi, O. Eltobgy, and M. Hefeeda, "Adaptive multicast streaming of virtual reality content to mobile users," in *Proc. Thematic Workshops ACM Multimedia*, 2017, pp. 170–178.

[3] Y. Sun, Z. Chen, M. Tao, and H. Liu, "Communication, computing and caching for mobile VR delivery: Modeling and trade-off," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2018, pp. 1–6.

[4] P. R. Alface, J. Macq, and N. Verzijp, "Interactive omnidirectional video delivery: A bandwidth-effective approach," *Bell Labs Tech. J.*, vol. 16, no. 4, pp. 135–147, Mar. 2012.

[5] L. Xie, Z. Xu, Y. Ban, X. Zhang, and Z. Guo, "360ProbDASH: Improving QoE of 360° video streaming using tile-based HTTP adaptive streaming," in *Proc. 25th ACM Int. Conf. Multimedia*, Mountain View, CA, USA, Oct. 2017, pp. 315–323.

[6] L. Sun, F. Duanmu, Y. Liu, Y. Wang, Y. Ye, H. Shi, and D. David, "Multi-path multi-tier 360° video streaming in 5G networks," in *Proc. 9th ACM Multimedia Syst. Conf.*, Amsterdam, The Netherlands, Jun. 2018, pp. 162–173.

[7] J. Park, J.-N. Hwang, and H.-Y. Wei, "Cross-layer optimization for VR video multicast systems," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2018, pp. 206–212.

[8] Y. Bao, T. Zhang, A. Pande, H. Wu, and X. Liu, "Motion-prediction-based multicast for 360° video transmissions," in *Proc. 14th Annu. IEEE Int. Conf. Sens., Commun., Netw. (SECON)*, San Diego, CA, USA, Jun. 2017, pp. 1–9.

[9] M.-H. Chen, K.-W. Hu, I.-H. Chung, and C.-F. Chou, "Towards VR/AR multimedia content multicast over wireless LAN," in *Proc. 16th IEEE Annu. Consum. Commun. Netw. Conf. (CCNC)*, Jan. 2019, pp. 1–6.

[10] C. Rosa, K. Pedersen, H. Wang, P.-H. Michaelsen, S. Barbera, E. Malkamaki, and B. Sébire, "Dual connectivity for LTE small cell evolution: Functionality and performance aspects," *IEEE Commun. Mag.*, vol. 54, no. 6, pp. 137–143, Jun. 2016.

[11] C. Pan, M. Elkashlan, J. Wang, J. Yuan, and L. Hanzo, "User-centric C-RAN architecture for ultra-dense 5G networks: Challenges and methodologies," *IEEE Commun. Mag.*, vol. 56, no. 6, pp. 14–20, Jun. 2018.

[12] A. Tzanakaki and M. Anastasopoulos, "Wireless-optical network convergence: Enabling the 5G architecture to support operational and end-user services," *IEEE Commun. Mag.*, vol. 55, no. 10, pp. 184–192, Oct. 2017.

[13] V. Sitzmann, A. Serrano, A. Pavel, M. Agrawala, D. Gutierrez, B. Masia, and G. Wetzstein, "Saliency in VR: How do people explore virtual environments?" *IEEE Trans. Vis. Comput. Graphics*, vol. 24, no. 4, pp. 1633–1642, Apr. 2018.

[14] Y. Ban, L. Xie, Z. Xu, X. Zhang, Z. Guo, and Y. Wang, "Cub360: Exploiting cross-users behaviors for viewport prediction in 360° video adaptive streaming," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2018, pp. 1–6.

[15] M. Zink, R. Sitaraman, and N. Klara, "Scalable 360° video stream delivery: Challenges, solutions, and opportunities," *Proc. IEEE*, vol. 107, no. 4, pp. 639–650, Apr. 2019.

[16] F. Qian, L. Ji, B. Han, and V. Gopalakrishnan, "Optimizing 360° video delivery over cellular networks," in *Proc. 5th Workshop Things Cellular, Oper., Appl. Challenges*, 2016, pp. 1–6.

[17] C. Perfecto, M. S. Elbamby, J. Del Ser, and M. Bennis, "Taming the latency in multi-user VR 360°: A QoE-aware deep learning-aided multicast framework," 2018, *arXiv:1811.07388*. [Online]. Available: https://arxiv.org/abs/1811.07388

[18] A. Prasad, M. A. Uusitalo, D. D. Navrátil, and M. Säily, "Challenges for enabling virtual reality broadcast using 5G small cell network," in *Proc. IEEE Wireless Commun. Netw. Conf. Workshops (WCNCW)*, Apr. 2018, pp. 220–225.

[19] A. Prasad, A. Maeder, and M. A. Uusitalo, "Optimizing over-the-air virtual reality broadcast transmissions with low-latency feedback," in *Proc. IEEE 5G World Forum (5GWF)*, Jul. 2018, pp. 112–117.

[20] C.-L. Fan, J. Lee, W. C. Lo, C.-Y. Huang, K.-T. Chen, and K.-T. Chen, "Fixation prediction for 360° video streaming in head-mounted virtual reality," in *Proc. 27th Workshop Netw. Oper. Syst. Support Digit. Audio Video*, 2017, pp. 67–72.

[21] M. Almquist, V. Almquist, V. Krishnamoorthi, N. Carlsson, and D. Eager, "The prefetch aggressiveness tradeoff in 360° video streaming," in *Proc. 9th ACM Multimedia Syst. Conf.*, 2018, pp. 258–269.

[22] M. S. Pan, T. M. Lin, C. Y. Chiu, and C. Y. Wang, "Downlink traffic scheduling for LTE-A small cell networks with dual connectivity enhancement," *IEEE Commun. Lett.*, vol. 20, no. 4, pp. 796–799, Apr. 2016.

[23] X. Zhang, R. Yu, Y. Zhang, Y. Gao, M. Im, L. G. Cuthbert, and W. Wang, "Energy-efficient multimedia transmissions through base station cooperation over heterogeneous cellular networks exploiting user behavior," *IEEE Wireless Commun.*, vol. 21, no. 4, pp. 54–61, Aug. 2014.

[24] P. Belotti, C. Kirches, S. Leyffer, J. Linderoth, J. Luedtke, and A. Mahajan, "Mixed-integer nonlinear optimization," *Acta Numerica*, vol. 22, pp. 1–131, May 2013.

[25] J. Yang, J. Luo, D. Meng, and J.-N. Hwang, "QoE-driven resource allocation optimized for uplink delivery of delay-sensitive VR video over cellular network," *IEEE Access*, vol. 7, pp. 60672–60683, 2019.

[26] J. D. Hedengren, R. A. Shishavan, K. M. Powell, and T. F. Edgar, "Nonlinear modeling, estimation and predictive control in APMonitor," *Comput. Chem. Eng.*, vol. 70, pp. 133–148, Nov. 2014.

[27] C. Wu, Z. Tan, Z. Wang, and S. Yang, "A dataset for exploring user behaviors in VR spherical video streaming," in *Proc. ACM Multimedia Syst. (MMSys)*, Taipei, Taiwan, Jun. 2017, pp. 193–198.
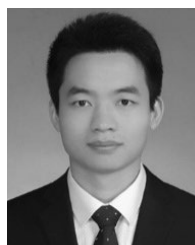
**JUNCHAO YANG** received the B.S. and M.S. degrees in communication and information engineering from the Chongqing University of Posts and Telecommunications, Chongqing, in 2008 and 2012, respectively, where he is currently pursuing the Ph.D. degree in communication and information engineering. His current research interests include multimedia networking, VR/AR video streaming, and convex optimization

**JIANGTAO LUO** received the B.S. degree from Nankai University, Tianjing, in 1993, and the Ph.D. degree from the Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Shanghai, in 1998.

From 1998 to 2002, he was an Associate Professor with the Key Laboratory of Laser Technology, Zhengzhou University. From 2002 to 2004, he was a Research Assistant with Shanghai Jiao Tong University. Since 2004, he has been the Chongqing University of Posts and Telecommunications, where he is currently the Deputy Dean of the Electronic Information and Networking Institute. He has authored one book, more than 150 articles, and more than 20 inventions. His research interests include big data, network traffic mining, and named date networking (NDN). He is the ACM member, since 2013.

**JUNXIA WANG** received M.S. degree in communication and information engineering from the Chongqing University of Posts and Telecommunications, Chongqing, in 2017, where she is currently pursuing the Ph.D. degree. Her current research interests include named date networks (NDN), network coding, and vehicle networks.
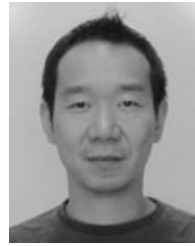
**SONG GUO** received the Ph.D. degree in computer science from the University of Ottawa. From 2007 to 2016, he was a Professor with The University of Aizu. He is currently a Full Professor with Department of Computing, The Hong Kong Polytechnic University. He published more than 400 articles in major conferences and journals. His research interests include big data, cloud computing and networking, and distributed systems. His work was recognized by the 2016 Annual Best of Computing: Notable Books and Articles in Computing in ACM Computing Reviews. He was a recipient of the 2017 IEEE Systems Journal Annual Best Paper Award and other five Best Paper Awards from IEEE/ACM conferences. His research has been sponsored by JSPS, JST, MIC, NSF, NSFC, and industrial companies. He also served as a General Chair, a TPC Chair, and a Symposium Chair for numerous IEEE conferences. He currently is the Member of Board of Governors and the Director of Membership Services of the IEEE Communications Society. He was an Associate Editor of the IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, from 2011 to 2015 and an IEEE Communications Society Distinguished Lecturer, from 2016 to 2017. He is currently on the Editorial Boards of the IEEE TRANSACTIONS ON CLOUD COMPUTING, the IEEE TRANSACTIONS ON EMERGING TOPICS IN COMPUTING, the IEEE TRANSACTIONS ON SUSTAINABLE COMPUTING, the IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING, and the IEEE COMMUNICATIONS.

• • •