

Research Article

Adaptive VR Video Data Transmission Method Using Mobile Edge Computing Based on AIoT Cloud VR

Min Wang ¹, Fuquan Zhang ^{2,3}, Linjuan Ma ⁴, and Ye Tian ⁵

¹Dean's office, Fujian Chuanzheng Communications College, Fuzhou, China 350007

²College of Computer and Control Engineering, Minjiang University, Fuzhou, China 350108

³Digital Media Art, Key Laboratory of Sichuan Province, Sichuan Conservatory of Music, Chengdu, China 610021

⁴School of Computer Science and Technology, Beijing Institute of Technology, Beijing, China 100081

⁵Experimental Art School, Sichuan Conservatory of Music, Chengdu, China 610041

Correspondence should be addressed to Fuquan Zhang; zfq@mju.edu.cn and Linjuan Ma; malinjuan@bit.edu.cn

Received 24 May 2022; Accepted 29 July 2022; Published 7 September 2022

Academic Editor: Saru Kumari

Copyright © 2022 Min Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Aiming at the high requirements of cloud service-based virtual reality in AIoT for data transmission rate and delay sensitivity, a cloud VR system scheme based on MEC (Mobile Edge Computing) is proposed, which mainly incorporates viewpoint-based VR video data processing and hybrid digital-to-analog (HDA) transmission optimization and can be served for AIoT transmission filed. Firstly, a learning-driven multiaccess MEC offloading strategy is designed, in which the VR terminal automatically selects the optimal MEC server for task offloading, thereby effectively improving network efficiency and reducing service delay. Secondly, the progressive transmission of the VR data is realized through viewpoint-aware dynamic streaming based on ROI (region of interest) and the priorities of different objects. The transmission priority of each object in the scene is determined through the ROI layering, which effectively solves the contradiction between the large data volume in the VR scenes and the network bandwidth limitation when applied in AIoT domain, and further improves the real-time performance of the system. Then, the HDA (hybrid digital-analog) technique is introduced to optimize the transmission. Finally, the base station protocol stack is modified on the basis of the LTE (Long-Term Evolution) system, and the MEC technology is integrated to realize a complete cloud VR system in AIoT. The experimental results show that compared with other advanced schemes, the proposed scheme can achieve more robust and efficient data transmission performance and provide better VR user experience.

1. Introduction

Recently, it is well known that VR (virtual reality) systems combined with AIoT, incorporating multiple technologies such as digital image processing, computer graphics, multimedia, computer simulation, sensors, and computer networks, have a wide range of applications in many fields such as entertainment, simulation training, aerospace, scientific and computing visualization, art, [1]. VR has three most prominent features: immersion, interactivity, and imagination, which are also known as the 3I characteristics of VR [2] and can realize human-computer interaction based on these characteristics. Cloud VR is a kind of real-time VR technology where can be used for amount data transmission in AIoT based on cloud computing, in which cloud servers

are used to replace users' local computing devices, which promotes the popularization of VR applications [3]. The cloud VR architecture consists of four parts: the content layer, the platform layer, the network layer, and the terminal layer. Among them, the content provided by the platform layer to the content layer includes cloud VR video services and cloud VR strong interaction services and is responsible for video import, transcoding, storage, broadcast control and distribution processing, logical calculation, and real-time rendering for strong interaction services. The network layer consists of four parts: backbone network, MAN (metropolitan area network), access network, and home network, which meet cloud VR's requirements for large bandwidth and low latency. Finally, the terminal layer is connected to the platform layer by accessing 5G/Wi-Fi to realize functions

such as VR content presentation and user authentication [4]. Cloud VR Solution Architecture is shown in Figure 1.

However, due to the huge amount of VR video data, in addition to cloud computing and rendering, network transmission bandwidth and delay limitations have become new bottlenecks for the entire system [5]. For basic 4 K resolution cloud VR services, the network bandwidth needs to reach at least 40 Mbit/s, and the RTT (round trip delay) of transmission should be controlled within 40 ms to provide users with a good viewing experience [6]. In the current mobile network architecture, the distance between the user and the server is at least at the metro distance level. Regardless of device forwarding and image transmission, the RTT of fiber transmission only is as high as 30–40 ms, which is difficult to meet the requirements of cloud VR [7].

With the development of 5G technology and the need of AIoT, the bandwidth of mobile networks has greatly increased. By sinking computing nodes to the vicinity of the application scenario (gateways), MEC puts the data collection and analysis operations close to the user side and realizes the closed loop of data processing on the edge, which can alleviate the network transmission pressure and shorten the data processing time [8]. Under the MEC architecture, deploying VR applications on MEC nodes will bring the following advantages: (1) the MEC computing nodes are directly deployed close to the mobile gateways, which can reduce the number of hops for network transmission between VR application data and end users, and reduce network processing delay [9]; (2) the VR application runs on the MEC node, which can realize local data processing. For the users under the same UPF (User Plane Function), the data does not need to enter the Internet, which can reduce the transmission pressure on the Internet [10]; (3) edge processing of VR content, for example, in VR live broadcast scenarios, the process of VR video splicing, encoding and transcoding, and distribution can be performed directly on the MEC nodes, data offloading can be realized for local users nearby, and OTT (Over the Top) users or users covered by other MEC nodes can be quickly distributed through CDN (content delivery network) [11]; and (4) the distributed networking of MEC can realize continuous VR experience in mobile scenarios, such as watching VR live broadcasts or participating in video conferences on high-speed mobile carriers (such as cars and high-speed trains) [12].

In addition, MEC can work with low-latency application-layer protocols such as QUIC (Quick User Datagram Protocol Internet Connection) and RTP (Real-Time Transport Protocol) to make cloud VR possible [13]. On the other hand, in the MEC scenario, the source ends (edge servers) are more closely connected with the channel ends (base stations), and the bandwidths are sufficient to support the transmission of baseband data between the servers and the base stations, which greatly improves the feasibility of using pseudoanalog, HAD (hybrid digital-analog) and other source-channel joint coding techniques [14].

When a user watches a VR video, due to the limitation of the FoV (field of view) of the display device, the user often can only watch a certain part of the whole video at a

moment. If the server transmits the entire video, most of the bandwidth resources will inevitably be wasted. Therefore, the adaptive block transmission method based on DASH (Dynamic Adaptive Streaming over HTTP) is the most widely used method for VR video streaming [15]. In the block transmission mode, a complete VR video is divided into many video blocks, and each video block is encoded into different quality levels. The server adaptively selects an optimal quality level for each video block according to factors such as network bandwidth and transmits it to the user. There are two basic schemes in the adaptive block transmission mode: view adaptation and rate adaptation [16]. The former is central for predicting changes in users' viewpoints, and the latter is central to resource allocation by the servers.

Based on the mobile edge computing technology in AIoT, this paper modified the LTE (Long-Term Evolution) base station protocol stack to build a mobile edge computing platform to serve for AIoT that expands to support HDA (hybrid digital-analog) transmission, in order to realize an efficient and reliable cloud VR system. The main contributions of this paper are as follows:

- (1) A learning-driven MEC server unloading strategy is adopted, so that users can automatically select the optimal MEC server
- (2) Realize a complete cloud VR system through viewpoint-aware dynamic streaming based on ROI and object priorities
- (3) Based on HDA technique, the system transmission efficiency is optimized to provide high quality VR videos under limited bandwidth

2. Related Research

The main content in the VR system is virtual scenes, and there will be a large number of 3D scenes in the virtual scenes. In the current network environment, the system is faced with the problem of how to solve the contradiction between the data transmission of 3D scenes and the limited network bandwidth in practice in AIoT. This aspect involves the processing, sending, and receiving of the 3D scenes, and at the same time, it is also necessary to ensure that the user-side scenes can be generated with a good visual experience [17]. Whether these problems can be solved are fundamentally related to the successful implementation of the VR system.

The researchers have proposed FOV transmission schemes for differential transmission of panoramic video information based on viewpoint areas [18]. These include the pyramid projection transmission scheme proposed by Facebook and the Tile Wise transmission scheme promoted by Huawei.

In the pyramid projection transmission scheme, a full-view nonuniform quality code stream is prepared for each view, and high-quality coding is used in the user's viewpoint region, while low-quality coding is used in other regions [19]. This method greatly reduces the bandwidth

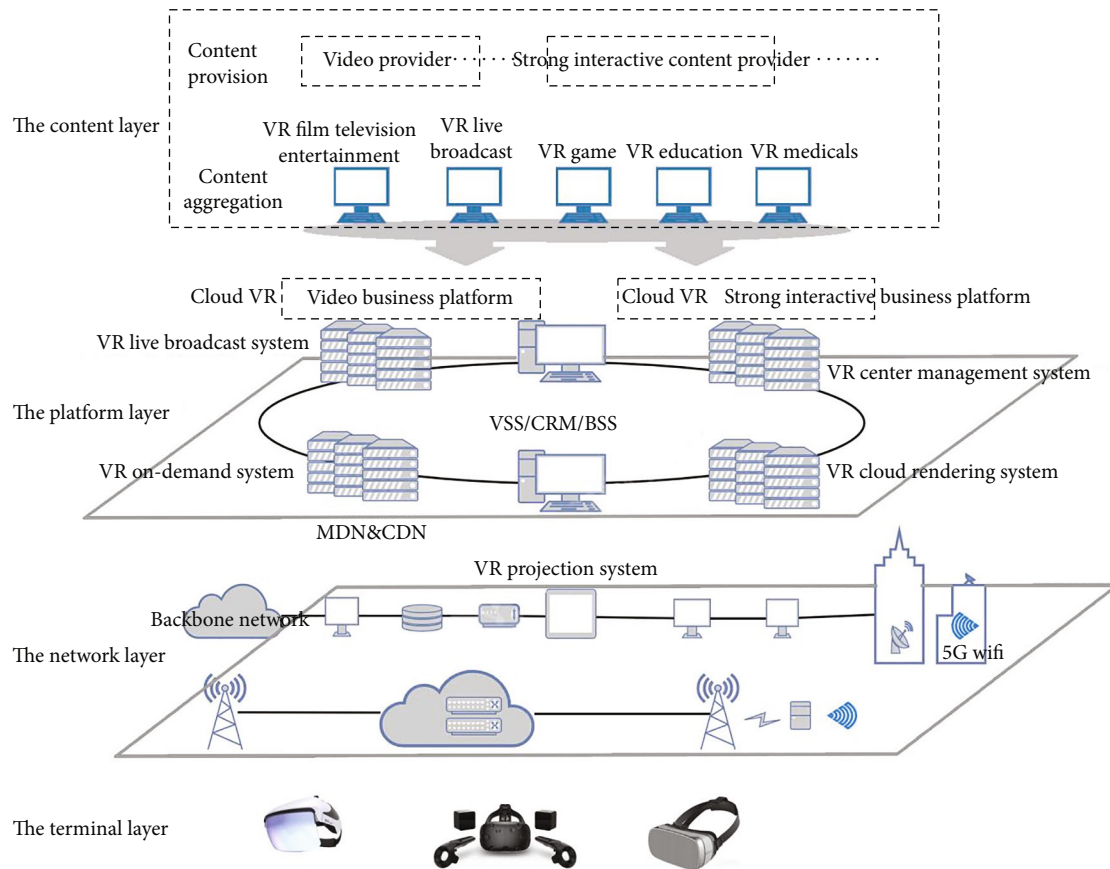


FIGURE 1: Cloud VR Solution Architecture.

requirements of system users when watching panoramic videos and improves the effective utilization of network bandwidth, but the sum of all perspective video files in the system server is more than 6 times that of the original files. The Tile Wise transmission scheme combines low-quality full-views and high-quality viewpoint regions. The server side does not need to prepare for each viewing angle area but divides the panoramic video image into multiple tiles at the same time, each area corresponds to a stream that can be decoded independently, and the server will prepare a low-quality panoramic full-view video stream. The client obtains a full-view stream and a high-quality tile selected according to the viewpoint information [20]. For the construction of panoramic images with nonuniform qualities, Hosseini et al. [21] proposed a viewpoint-aware adaptive VR transmission framework based on extended MPEG-DASH SRD (moving picture experts group-dynamic adaptive streaming over HTTP spatial relation description). Similarly, Kim et al. [22] proposed a SSAS (spatial segmented adaptive streaming) scheme based on the HLS (HTTP live streaming) protocol to realize real-time adaptive streaming based on user viewpoints. These solutions draw on existing HTTP adaptive transmission protocols such as DASH and further expand time-based tiling to space to achieve dynamic adaptive streaming.

According to the situational information such as the computing capability of the users' mobile terminals, appropriate MEC servers are selected for efficient task offloading,

so as to ensure the network delay performance and reduce energy consumption. Guo et al. [23] proposed a MEC task offloading strategy based on nonorthogonal multiple access, which takes into account the constraints of different access technologies. By considering different business quality of service (QoS) constraints, Henri et al. [24] proposed an offloading strategy that can guarantee a strong delay boundary based on game theory. Based on the Stackelberg game theory Hosseini et al. [25] proposed a price-based distributed MEC task offloading algorithm, which enables users to make autonomous decisions. In addition, Liu and Liu [26] proposed an energy-efficient MEC task offloading algorithm for ultradense wireless network scenarios in which energy overheads are minimized by optimizing offloading decision variables and power bandwidth allocation. In the existing research on MEC task offloading, it is assumed that the computing power and storage capacity of the MEC are known, and based on the research scenario of a single MEC server, the offloading decision of computing tasks is made with the goal of optimal delay or optimal energy. However, with the densification of base station deployment in 5G networks, a large number of MEC servers will be deployed on base stations or access points (APs) that are closer to user mobile terminals. The computing and storage capabilities of different MEC servers are different. Therefore, the mobile terminals on the user side need to independently decide and select the optimal MEC server access strategy according to

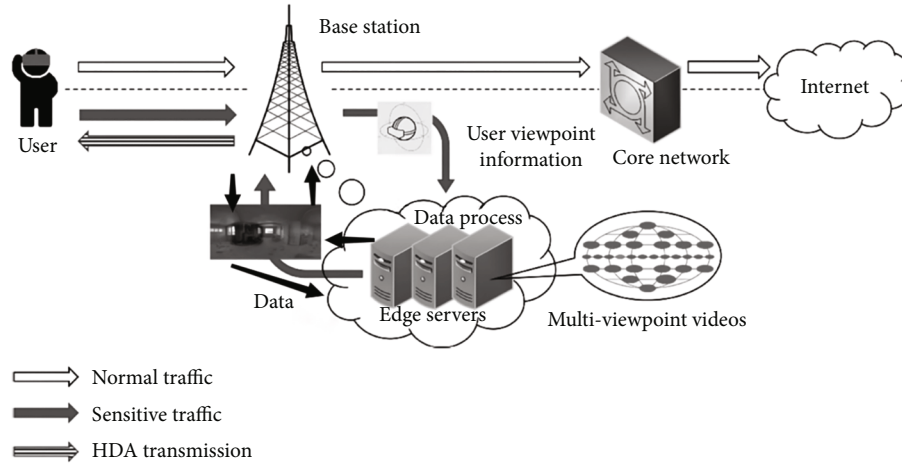


FIGURE 2: System structure.

the situational information such as service characteristics and network environment, so as to minimize network delay and network energy consumption at the same time, thereby realizing an energy-efficient MEC server task offloading strategy.

Liu et al. [27] proposed an efficient VR transmission mechanism based on source-channel joint coding. After tiling the VR videos with reference to the users' FOV information, different levels of error correction strategies are used to maximize the viewing quality within users' FOV. Feng et al. [28] defined a new QoE (Quality of Experience) metric to measure the user's viewing experience and presented an efficient modulation control algorithm to maximize the QoE value under different channel conditions. Zhang and Ma [32] proposed multiobject crowd real-time tracking in the dynamic environment based on a novel neural network, which can be used in AIoT cloud VR.

3. VR Adaptive Transmission Scheme Based on MEC and Viewpoint Awareness

3.1. System Architecture. The proposed solution integrates the MEC technology on the basis of the LTE system and expands the base station to realize the HDA transmission mode, so as to meet the high requirements of the interactive VR services for latency and network quality and provide high-performance network support for the cloud VR service especially in AIoT data transmission [32]. The system structure is shown in Figure 2.

By modifying the protocol stack of the base station, while introducing the MEC function, the compatibility of the system to standard LTE terminals is maintained. The tunneling protocols are used to redirect traffic at the network layer to filter and offload sensitive traffic from the edge services. The base station maintains a sensitive traffic table of the edge services, which records the IP addresses, protocols, and port numbers of data packets that need to be forwarded to the edge servers. Each passing data packet is matched. If the data packet matches the entry in the sensitive table, the destina-

tion IP in the tunnel packet header of the GPRS Tunneling Protocol (GTP) is reconstructed, and the original core network IP is replaced with the edge server IP; that is, the data packet is forwarded to the edge server. For the returned downlink data, the edge server masquerades the source IP address as the real public network address of the application server.

The proposed architecture implements cloud VR based on the MEC system. The computing tasks and services are moved down to the edge of the base stations, so as to minimize the transmission delays from both the network structures and the physical distances, and the RTT will be controlled within 10 ms. It not only improves the response speed of the server but also ensures the stability of the network service quality and greatly improves the users' viewing experience. The introduction of HDA technique provides a more flexible transmission mode for edge servers, enabling them to make full use of bandwidth resources and alleviating the saturation effect of existing digital transmission in AIoT.

3.2. Learning-Driven MEC Server Adaptive Offloading Strategy. In the MAB (multiarmed bandit) model [24], there are N gambling arms and one player for multiple rounds of selection. Each time the player selects one of the gambling arms and receives the corresponding reward, the player can only obtain the reward value of the selected arm after selection. The reward value of each gambling arm follows some unknown specific distribution and is independent of each other. The player learns the reward distribution of different gambling arms through exploration and utilization. After J rounds of games, the optimization goal of the player is to maximize the expected value of the reward.

It is assumed that there are U users and M base stations in the 5G wireless network scenario, and each base station contains an MEC server (to simplify the description, the base station and the MEC server are collectively represented by M). Let the total system bandwidth be B , and there are K subcarriers in the system bandwidth. Assuming that a user can only access one base station at time t , and at most, one user can access a subcarrier, then we have

$$\sum_{k \in K, m \in M} a^{k,j,m}(t) = 1, \forall i = U, \quad (1)$$

where $k \in K$ denotes a resource block. The SINR (signal to interference plus noise ratio) of user terminal i and base station m on resource block k is

$$\Gamma_i^{k,m} = \frac{P_{im}^{k,m} g^{k,i,m}}{N_0}, \quad (2)$$

where $P_{im}^{k,m}$ represents the transmission power from base station m to user i in resource block k , the channel gain between base station m and user i is $g^{k,i,m}$, and N_0 is the noise power that follows $N(0, \delta)$ distribution. The transmission rate from the user to the base station is

$$R_i^m = \sum_{k_m=1}^{K_m} a_{i_m}^{k_m} \log \left(1 + \Gamma_i^{k,m} \right), k_m \in K_M^B. \quad (3)$$

For delay-sensitive services, it is assumed that the arrival rate of the data packets conforms to the Poisson distribution with the arrival rate of λ_{ds} , and the fixed length of the data packets is L_{ds} . In order to meet the QoS constraints of the delay-sensitive services, based on the effective bandwidth theory, the effective bandwidth with the transmission delay bound is defined as

$$W(\theta_v) = \lim_{t \rightarrow \infty} \frac{1}{t\theta_v} \log E \left(e^{\theta_v Z(t)} \right), \quad (4)$$

where $W(\theta_v)$ is the effective bandwidth, θ_v is the QoS value of the user terminal, $Z(t)$ is the number of packets reached within the period $(0, t)$, and $E(\cdot)$ represents the mathematical expectation.

Assuming that the maximum computing frequency of each MEC server (base station) is $f_m^{\max} (\forall m \in M)$, $f(t) = [f_i, m(t)]$, $m \in M$, and $i \in U$, the amount of computation that the MEC server m can provide for the user's mobile terminals can be expressed as

$$A_{i,m} = \frac{f_{i,m} T}{b_i}, \quad (5)$$

where $f_{i,m}$ represents the computing frequency of each server, and b_i represents the computing load of the user-side tasks, which can be measured in an offline fashion. The MEC network architecture is shown in Figure 3.

In the proposed learning-driven MEC-MAB autonomous offloading algorithm, the user's mobile terminal i is the player, and the MEC server m is the gambling arm. If the user i chooses to access the MEC server m , the corresponding random reward value $Q_{i,m}$ will be obtained. The reward value of each MEC server obeys a specific distribution with mean value as $\pi = [\pi_1, \pi_2, \dots, \pi_m]$ and is independent of each other, where π_m is the real reward of MEC server m . Since the user cannot always choose the server with the highest real reward, the regret value R_j is defined

as the difference between the actual reward value obtained after j selections and the expected maximum reward value:

$$R_j = \pi^* j - \sum_{m=1}^M E[N_j(m)] \pi_m, \quad (6)$$

where $\pi^* = \max_{1 \leq s \leq M} \pi_s$, and $N_j(m)$ is the number of times the MEC server m has been selected in the previous j rounds. Since in the MAB model, the real reward value of the gambling arm is the reward value generated after the action is performed, it is necessary to estimate the reward value for the selection behavior of the gambling arm as follows:

$$W_{j+1}(m) = W_j(m) = \frac{1}{N_j(m)} \left[r_{N_j(m)} - W_j(m) \right]. \quad (7)$$

Using the Thompson-Sampling algorithm [29], the probability of each selection of the reward value of the MEC server in the MAB model is regarded as a Beta (α, β) distribution, and then the reward value distribution probability function of the MEC server selection behavior can be mathematically expressed as

$$f(m, \alpha, \beta) = \frac{1}{B(\alpha, \beta)} m^{\alpha-1} (1-m)^{\beta-1}. \quad (8)$$

The parameter update rule for Beta distribution is given as

$$(\alpha_1, \beta_1) = \begin{cases} (\alpha_1, \beta_1) f_i, & f_i \neq l, \\ (\alpha_1, \beta_1 + (r_l, 1 - r_l)) f_i, & f_i = l. \end{cases} \quad (9)$$

Initially, the user mobile terminal observes situational information such as QoS of its computing task and sets $t = 0$ and $\gamma = 0$. at time t , $t \leq T$, the reward estimation of the user's mobile terminal for the MEC server selection behavior satisfies $W(m) \sim \text{Beta}(\alpha_m, \beta_m)$. The user selects the MEC server with the largest reward value $\arg \max_m W(m) \rightarrow \text{MEC}_i$. The network applies the selected access behavior and measures the corresponding reward value r_t , and the parameters are updated as $(\alpha_1, \beta_1) + (r_t, 1 - r_t) \rightarrow (\alpha_1, \beta_1)$.

In the proposed MEC-MAB algorithm, as the number of observations from the MEC server selections increases, the confidence interval of the beta distribution becomes narrower, enabling the users to automatically select the optimal MEC server with maximum reward.

3.3. Viewpoint-Aware Progressive Dynamic Streaming. This paper proposes a viewpoint-aware dynamic streaming based on ROI and object priority. The strategy first divides the data into the current scenes, potential scenes, and future scenes based on ROI judgment. And the concept of ROI is extended, and the priorities of vertical and horizontal objects are determined through ROI layering, so as to determine the transmission priority of each object in the scene.

The visibility judgment and elimination of object space can reduce the transmission of unnecessary scenes in order

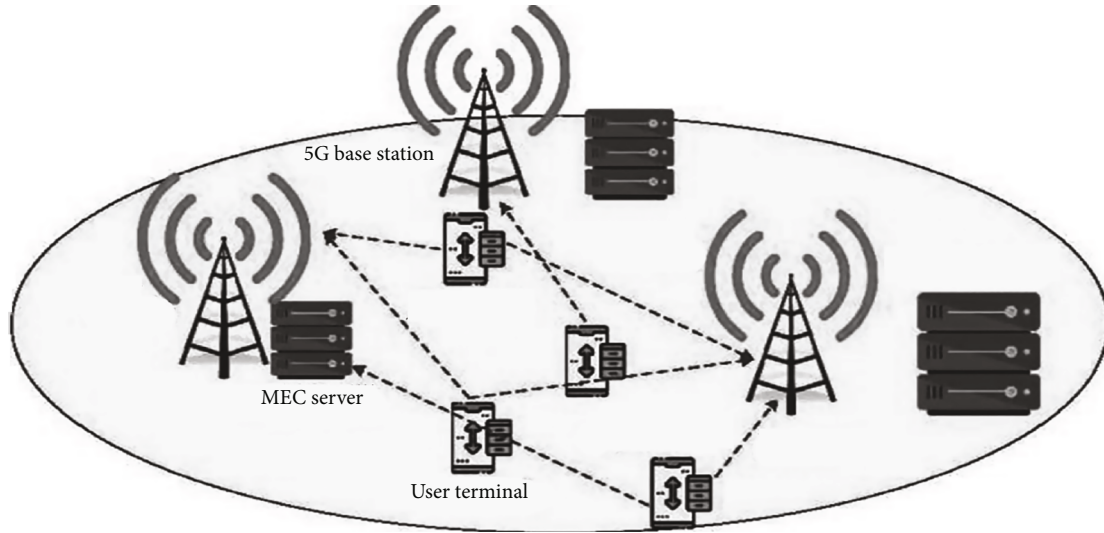


FIGURE 3: MEC network architect.

to achieve the minimum transmission volume, thereby improving the real-time performance of interaction. To achieve this goal, the whole scene must first be analyzed to determine the visible scenes according to the position and angle of the user, and the visibility should be calculated according to the relevant algorithm to remove those unnecessary or unimportant scenes. Then, the current visible scenes are firstly transmitted, and as the viewpoint moves, the incremental part of the scenes is gradually transmitted, which is also the original intention of progressive transmission.

As shown in Figure 4, the viewpoint range is divided into three areas: CPVS (Current Potential Visible Scenes), IPVS (Incremental Potential Visible Scenes), and FPVS (Future Potential Visible Scenes), based on the visual habit of looking directly in front and then looking around, and observing the near area first and then the far area.

CPVS Zone is the immediate and nearest currently visible scene area. All object models in this area are visible to the user and are relatively close to the user; so, objects in this area should have the highest priority in the process of transmission and interaction. IPVS Zone is as follows: this zone consists of two parts. To the user, the objects in this area are not immediately visible and are relatively far from the user. If the user's viewpoint moves (walks forward or turns), then the objects in this area are immediately visible. Therefore, the object model in this area should have a relatively high priority. After the object model in the CPVS Zone is accessed, the scenes in this area are downloaded first, so that these models can be displayed in time when the user roams. FPVS Zone is as follows: all object models in this area are not currently visible to the user and are relatively far away from the user. Object models in this area should have lower priority, and these objects can be prefetched to the client for use in scene roaming only when the network is idle or if additional network bandwidth is available. Thus, the visual areas are divided into CPVS, IPVS, and FPVS. The object models in this three zones correspond to three queues, queue

1, queue 2, and queue 3, respectively, and the priority of the three queues is $\text{queue 1} > \text{queue 2} > \text{queue 3}$.

When there are many objects in the scene, in order to better distinguish different object models and realize the progressive transmission of the scene, it is necessary to further determine the access sequence of a certain scene with multiple objects in it; that is, it is necessary to determine the visual importance of different objects in a certain area. This paper extends the concept of ROI and determines the order of a specific object in the access queue by layering a certain ROI area and considering the horizontal and vertical importance of different objects.

Level of ROI is from the servers' perspective, the ROI area of the current user is further subdivided into several levels, and the transmission order of the objects is determined accordingly. As shown in Figure 5, if there are multiple objects in the currently visible ROI, the distance and viewing angle can be used to determine the order in which the objects are accessed: objects in level A have the highest priority, objects in level B have the next priority, and objects in level C have the lowest priority.

In this way, the system mainly uses limited bandwidth resources to transmit videos within the user's visible range, compresses redundant content as much as possible, and provides the best viewing effect. At the same time, each frame of video transmitted contains full-view information, which can achieve "device-cloud asynchronous" rendering. When the user's posture changes, the local display device does not need to wait for the server to send back data and completes the rendering locally in real time, updating the scene with the shortest delay to ensure the complete and smooth transition.

The essence of progressive streaming is "download while browsing" to achieve the optimal real-time effect. According to the principle of human vision, the closer the object is to the viewpoint, the smaller the angle that the object deviates from the viewpoint, and the higher the resolution of the object that the viewpoint can observe. To save bandwidth, it is not necessary to download the full model increments,

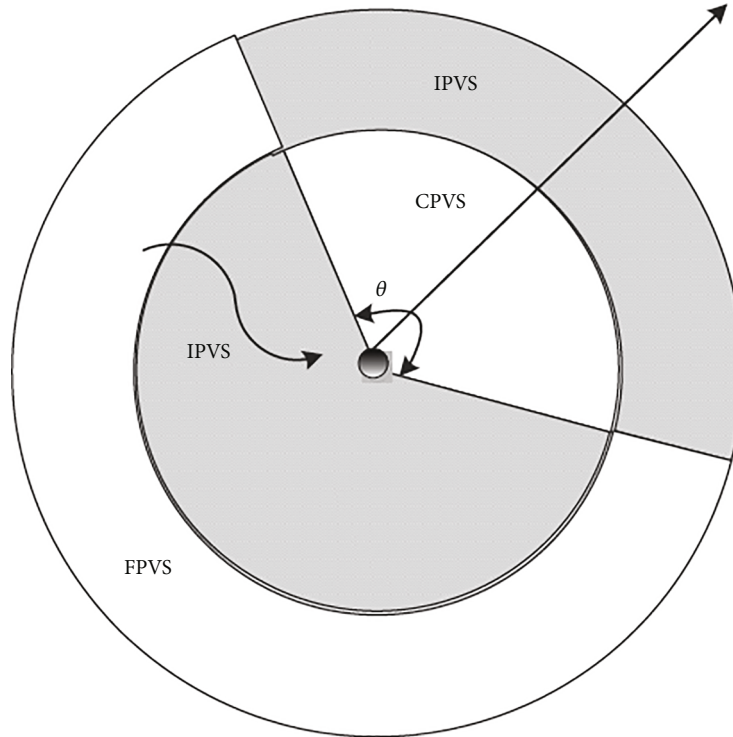


FIGURE 4: ROI region of the user.

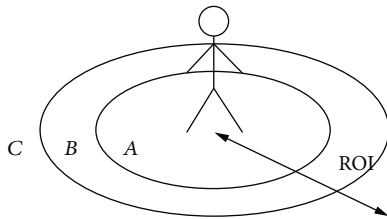


FIGURE 5: ROI layering.

instead, we can just download the ORM of the scene. The ORM of an object can be determined based on its visual importance to the viewpoint:

$$W(O_i) = \left(1 - \frac{D_i}{R}\right) \cos \frac{\theta_i}{2}, \quad D_i < R, \quad (10)$$

where $W(O_i)$ represents the visual importance of the object O_i , R is the ROI radius of the avatar, D_i is the distance between O_i and the avatar's viewpoint, and θ_i represents the deviation angle between the object i and the avatar's viewpoint ($0 \leq \theta \leq 180^\circ$).

3.4. Optimized HDA Transmission Scheme. In the existing wireless communication system, the channel end encodes the video data into a bit stream for transmission. If there is a bit error in the transmission of the video code stream, the decoding of the video data will cause serious visual distortion or even a decoding failure. Although the current wireless video soft transmission scheme can realize seamless adaptation of video transmission quality to channel conditions, its transmission efficiency is not satisfactory. Combin-

ing the high efficiency of traditional digital transmission and the robustness of video soft transmission, the HDA transmission technique has the potential to provide stable, reliable, and high-efficiency VR video transmission in AIoT.

In the proposed HDA transmission system, a time-division multiplexing HDA video soft transmission scheme is designed. The video in the user window is decomposed into two layers: the first layer is the base layer signal, which is generated from the video source compressed by the HEVC encoder. The second layer is the enhancement layer signal, which is the residual value after subtracting the original video signal and the reconstructed signal of the first layer. The two layers of video signals are transmitted in a time-division multiplexing manner. On the one hand, in order to achieve reliable transmission of the digital part, the target bit rate is controlled by the quantization parameter, and the channel coding rate and modulation order are determined by the SNR. On the other hand, overall video quality is directly dependent on the MSE (mean squared error) of the analog signal, which can be expressed as a function of the data variance of the analog part, the power and bandwidth allocated to the analog part, and the noise power of the channel.

In terms of power allocation, it is first necessary to ensure that the base layer can be successfully decoded. Therefore, the overall video quality of HDA video transmission is determined by the data variance of the enhancement layer (analog source), the resources allocated to the analog part, and the channel noise power. Based on the SNR, the bitstream signals at the base layer are turbo-coded with a selected channel coding rate, and the coded signals are subjected to quadrature amplitude modulation. Considering

that HEVC (high efficiency video coding) has basically removed the interframe correlation of video sequences, the residual between the original video and the reconstructed video basically does not contain interframe redundancy. The residual part is further decorrelated by 3D-DCT transform, and the power-scaled DCT coefficients are used to modulate the signal amplitude.

Since the signals in the second layer are the enhancement signal of the video, under the condition of limited bandwidth and power, restoring as many of the enhancement layer signals as possible helps to improve the quality of the reconstructed video. After the decorrelation operation of the enhancement layer signals, the energy distribution of the analog coefficients is relatively concentrated, which is manifested in some large coefficients that are concentrated in the upper left corner. In time-division multiplexing coding, appropriate parameters should be selected for the code rate and channel coding modulation mode of the first layer to ensure the correct decoding. Since the first layer is designed to be decoded correctly at a given channel noise power, the overall system distortion is determined by the reconstruction distortion of the second layer. In order to reduce the interference of the large coefficients of the analog part to the digital signal, we try to transmit the large coefficients by time-division multiplexing. Due to the bandwidth limitation, the small coefficients will be discarded. Although discarding small coefficients saves bandwidth, the high-frequency component information carried by these small coefficients cannot be recovered at the receiving end, which will bring additional performance loss.

Without loss of generality, we use MSE as the distortion measure. Let D_a and D_d be analog distortion and digital distortion, respectively. In order to successfully decode the digital base layer, the SNR of the digital part must be greater than the signal-to-noise ratio threshold SNR_{th} . The spectral efficiency corresponding to SNR_{th} is e_f , which depends on the MCS (modulation and coding scheme), and it must satisfy

$$\text{SNR}_{\text{th}} \leq \frac{P_d}{\sigma_n^2}, \quad (11)$$

where P_d is the average power distribution coefficient of the base layer digital signal, and σ_n^2 is the channel noise power with Gaussian white noise added. Since the sum of digital power and analog power is limited by the total power P_T , the power constraint relationship can be expressed as

$$P_T = P_a B_a + P_d B_d, \quad (12)$$

where P_a is the average power allocated to the enhancement layer analog signals, B_a is the bandwidth occupied by the transmitted analog signal, and B_d is the bandwidth occupied by the transmitted digital signal. After the enhancement layer is transformed by 3D-DCT, each group of video frames is further divided into N blocks, and the variance of the i -th block is defined as λ_i . The analog sig-

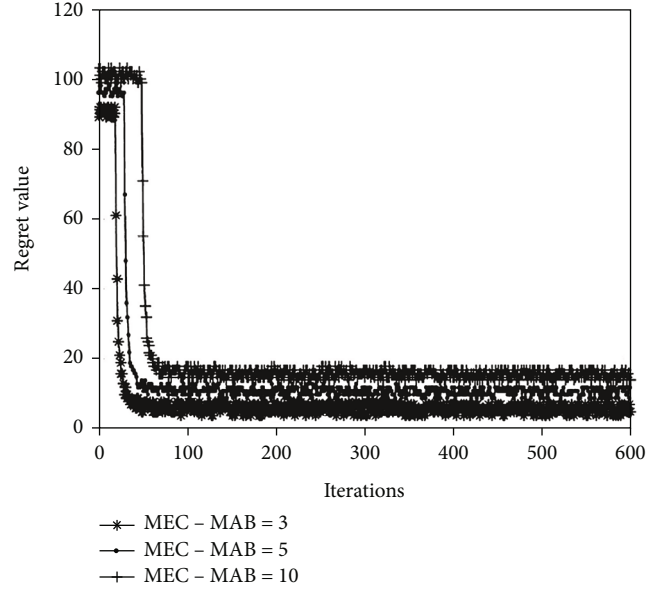


FIGURE 6: Validation of the proposed MEC offloading strategy.

nals transmitted through the channel are interfered by the channel noise, and the distortions can be expressed as [30]

$$D_{a1} = \sigma_n^2 \left[\left(\sum_{i=1}^{B_a} \sqrt{\frac{\lambda_i}{B_a P_a}} \right)^2 \right]. \quad (13)$$

On the other hand, the discarded coefficients cannot be recovered at the receiver, which also introduces additional distortion D_{a2} :

$$D_{a2} = \sum_{i=B_a+1}^N \lambda_i. \quad (14)$$

Therefore, the optimal power allocation problem can be defined as

$$\begin{cases} \min_{P_a, QP} D_a \\ \text{STB}_a P_a + B_d P_d = P_T, \\ B_a + B_d = N. \end{cases} \quad (15)$$

The variance of the i -th block can be expressed as [31]

$$\lambda_i = e^{k_i QP + w_i}, \quad (16)$$

where QP is the quantization parameter, and k_i and w_i are two parameters in the i -th block that represent the exponential relationship between λ_i and QP. After the video is digitally compressed and encoded, the number of quantized bits per pixel can also be further fitted with an exponential function, and the fitting parameters are a and b , respectively.

TABLE 1: PSNR results (dB) of HEVC scheme and proposed HDA scheme under different QPs.

Methods	QP = 20			QP = 25			
	$\beta = 0.8$	$\beta = 0.7$	$\beta = 0.5$	$\beta = 0.8$	$\beta = 0.7$	$\beta = 0.5$	$\beta = 0.25$
HEVC	46	46	46	42	42	42	42
Proposed scheme	53	52	51	50	48	46	41

The relationship between the quantization parameter QP and the number of bits produced per pixel R is

$$R(QP) = ae^{bQP}. \quad (17)$$

When a group of video frames has M pixels, the total number of bits obtained after digital compression can be calculated as

$$\text{Bit}(QP) = R(QP)M = ae^{bQP}M. \quad (18)$$

4. Experiment and Results

Based on the software radio platform from the laboratory and the modified LTE protocol stack of the MEC architecture, a complete MEC platform is build, and it is used as the bearer network to develop the cloud VR system which is suitable for AIoT. The system can use standard commercial terminals or professional VR headsets as client display devices. A complete performance evaluation of the proposed system is performed on this platform.

4.1. MEC Offloading Algorithm Verification. Firstly, the proposed learning-driven MEC task offloading strategy is verified by simulation. Assuming that the number of users is 10, the computing tasks of the users' mobile terminals follow the Poisson distribution, and the path loss exponent is set to 2.

The variation of the simulation regret value with respect to the number of iterations for the number of MEC server nodes (base stations) of 3, 5, and 10 is shown in Figure 6. It can be observed that the network regret value converges in a short time for different numbers of MEC servers. As the number of MEC servers increases, the convergence speed of the algorithm becomes slower, but the overall convergence speed is still reasonable, which shows that the proposed MEC-MAB offloading strategy has good convergence performance.

4.2. VR Transmission Scheme Validation. MATLAB experimental simulation of the proposed HDA transmission scheme is carried out. The HDA transmission system proposed in this paper consists of a data channel and a control channel. The data channel executes the function blocks of the transmit ends and receive ends, respectively, and is based on the power distribution calculation to solve P_a and QP with minimized distortions. The data channel combines digital transmission and pseudoanalog transmission. The digital transmission scheme uses HEVC for source coding and the LTE-based adaptive modulation and coding scheme for transmission. Different combinations of channel coding rates and modulation modes can be selected.

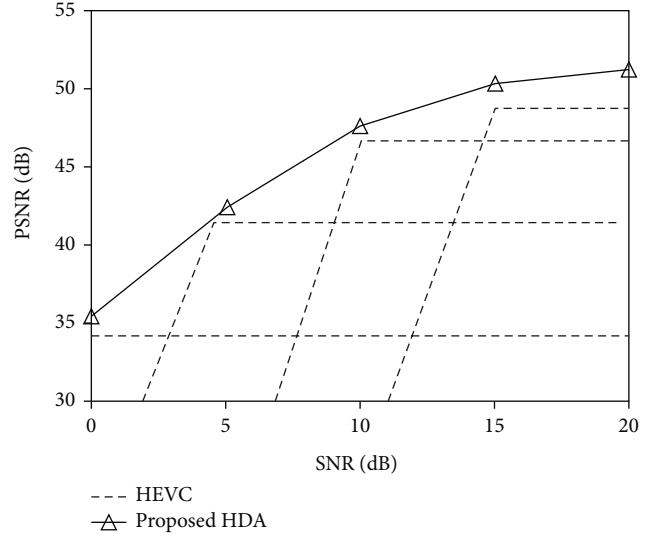


FIGURE 7: PSNR performance comparison.

Experimental simulations were performed using standard HD sequences with a resolution of 1664×1664 pixels. For analog transmission, each frame in the video sequence is divided into 64 blocks. In the experiment, each picture group is set to consist of 16 frames of images; so, the analog symbols of each picture group are divided into 1024 coefficient blocks. When the video frame rate is 30 fps, the source bandwidth N_s is 41.5 MHz. The bandwidth used for data transmission is defined as N_c , and a specific implementation process is designed to make the number of symbols used for source coding and channel coding in the digital part less than or equal to N_c .

The performance of the HDA transmission scheme proposed in this paper is compared with the existing digital video transmission scheme HEVC. PSNR is measured at the receiving ends to evaluate the quality of video transmission, and same bandwidth and power are used for the two schemes. According to the LTE adaptive modulation and coding scheme, the channel coding adopts LTE turbo coding, and the code rate is $R = 1/3$. The modulation scheme supports QPSK, 16QAM, and 64QAM. Taking the channel SNR = 5.5 dB as an example, the spectral efficiency is about 1.47. Table 1 gives the results of the HEVC scheme and the proposed HDA scheme on the test sequence under different QPs when the target channel SNR = 10 dB and the spectral efficiency of the digital part is 1.47. The ratio of the available video channel bandwidth to the source bandwidth is set to β , i.e., $\beta = N_c$.

It can be seen from Table 1 that when the digital part adopts a certain QP value, the video quality of the receiving

end under the HEVC scheme does not change with the change of the available channel bandwidth. When QP = 20, if the available channel bandwidth is only 1/4 of the source bandwidth, the data length of the digital part will exceed the available bandwidth, causing the digital part to not be decoded correctly. In contrast, with the proposed method, under the condition that the available bandwidth resources are severely limited, increasing the QP can realize the encoding and transmission of the digital part of the data.

Next, further performance comparisons between the proposed HDA scheme and HEVC scheme are carried out. Under different channel conditions, the classic HEVC scheme is inevitably affected by the cliff effect. As the SNR increases, so does the spectral efficiency, at which point HEVC will have the opportunity to choose a lower QP. Consider an SNR of 0 to 20 dB, $\beta = 0.5$, and 10% of the bandwidth is reserved for hybrid automatic retransmission of the digital part. As shown in Figure 7, the average PSNR of the proposed HDA scheme is 0.41 dB higher than that of the HEVC scheme. By adding analog signals to the existing digital transmission scheme and dividing part of the bandwidth to the analog signals, the saturation effect of the video quality at the receiving ends can be improved. When the SNR of the target channel is high, analog signal transmission can further achieve greater performance gains and benefit for AIoT.

5. Conclusions

With the rapid development of multimedia services such as VR, there are great requirements for the data transmission in AIoT; also, the resolution of video data is increasing day by day, along with the growing challenge for network bandwidth. Therefore, new business models, such as cloud VR, attracted more attention on the transmission network latency. In this paper, we design and build a cloud VR system based on AIoT using MEC technology to perform dynamic streaming relying on user viewpoint information. Through the learning-driven MEC autonomous offloading strategy, in the absence of prior information such as MEC server computing and storage capabilities and channel status, the optimal MEC server is autonomously selected for task offloading, and the energy consumption is minimized while satisfying user delay constraints. Combined with the viewpoint-aware progressive streaming based on ROI and object priorities used on the server side, the bandwidth requirements are reduced, and a good VR viewing experience is achieved. Meanwhile, based on the relationship between edge servers and base stations in the MEC architecture, HDA transmission technology is merged to further optimize the transmission bandwidth and efficiency in AIoT cloud VR systems. It is known that the use of the AIoT data transmission can enrich the data information and enhance the human-computer interaction in AIoT cloud VR systems. However, it needs more rapidly data processing and more faster network support. In subsequent studies, the availability of better compression schemes for panoramic video and AIoT data transmission in AIoT cloud VR will be further

explored, and the channel fading will be further considered with a view to achieving more performance gains.

Data Availability

The data we used is available, and the performance optimization scheme proposed in this paper can be used in adaptive data transmission of VR videos. And part of them are available from the corresponding authors upon request (malinjuan@bit.edu.cn; zfq@mju.edu.cn).

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the study of this work and publication of this paper.

Acknowledgments

Our work is supported by the Digital Media Art, Key Laboratory of Sichuan Province, Sichuan Conservatory of Music (Project No. 21DMAKL01), First Batch of Industry-University Cooperation Collaborative Education Project funded by the Ministry of Education of the People's Republic of China, 2021 (Project No. 202101071001), Minjiang University School-Level Scientific Research Project Funding (Project Nos. MYK17021, MYK18033, and MYK21011), and Minjiang University Introduced Talents Scientific Research Start-Up Fund (Project No. MJY21030).

References

- [1] P. Halarnkar, S. Shah, H. Shah, H. Shah, and A. Shah, "A review on virtual reality," *International Journal of Computer Science Issues (IJCSI)*, vol. 9, no. 6, p. 325, 2012.
- [2] Z. Lv, "Virtual reality in the context of Internet of Things," *Neural Computing and Applications*, vol. 32, no. 13, pp. 9593–9602, 2020.
- [3] T. Kämäräinen, M. Siekkinen, J. Eerikäinen, and A. Ylä-Jääski, "CloudVR: cloud accelerated interactive mobile virtual reality," in *Proceedings of the 26th ACM international conference on Multimedia*, pp. 1181–1189, 2018.
- [4] Z. Guo, P. Zhang, and J. Xia, "Design of virtual reality education platform based on 5G MEC," in *2021 20th International Conference on Ubiquitous Computing and Communications (IUCC/CIT/DSCI/SmartCNS)*, pp. 572–578, 2021.
- [5] A. Mehrabi, M. Siekkinen, T. Kämäräinen, and A. Ylä-Jääski, "Multi-tier cloudVR," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 17, no. 2, pp. 1–24, 2021.
- [6] X. Tang, C. Cao, Y. Wang et al., "Computing power network: the architecture of convergence of computing and networking towards 6G requirement," *China Communications*, vol. 18, no. 2, pp. 175–185, 2021.
- [7] Q. Cheng, H. Shan, W. Zhuang, L. Yu, Z. Zhang, and T. Q. Quek, "Design and analysis of MEC- and proactive caching-based 360° mobile VR video streaming," *IEEE Transactions on Multimedia*, vol. 24, pp. 1529–1544, 2022.
- [8] Y. Siriwardhana, P. Porambage, M. Liyanage, and M. Ylianttila, "A survey on mobile augmented reality with 5G mobile edge computing: architectures, applications, and

- technical aspects,” *IEEE Communications Surveys & Tutorials*, vol. 23, no. 2, pp. 1160–1192, 2021.
- [9] K. Bilal and A. Erbad, “Edge computing for interactive media and video streaming,” in *Second International Conference on Fog and Mobile Edge Computing (FMEC)*, pp. 68–73, 2017.
- [10] R. K. Srinivasa, N. K. S. Naidu, S. Maheshwari, C. Bharathi, and A. H. Kumar, “Minimizing latency for 5G multimedia and V2X applications using mobile edge computing,” in *2019 2nd International Conference on Intelligent Communication and Computational Techniques (ICCT)*, pp. 213–217, 2019.
- [11] J. Gedeon, F. Brandherm, R. Egert, T. Grube, and M. Muhlhauser, “What the fog? Edge computing revisited: promises, applications and future challenges,” *IEEE Access*, vol. 7, pp. 152847–152878, 2019.
- [12] L. Wang, L. Jiao, T. He, J. Li, and M. Mühlhäuser, “Service entity placement for social virtual reality applications in edge computing,” in *IEEE INFOCOM 2018-IEEE Conference on Computer Communications*, pp. 468–476, 2018.
- [13] A. Bujari, O. Gaggi, M. Luglio et al., “Addressing the bandwidth demand of immersive applications through NFV in a 5G network,” *Mobile Networks and Applications*, vol. 25, no. 3, pp. 1114–1121, 2020.
- [14] D. He, C. Luo, F. Wu, and W. Zeng, “Swift: a hybrid digital-analog scheme for low-delay transmission of mobile stereo video,” in *Proceedings of the 18th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, pp. 327–336, 2015.
- [15] D. Podborski, E. Thomas, M. M. Hannuksela, S. Oh, T. Stockhammer, and S. Pham, “Virtual reality and DASH,” in *International broadcasting convention, Ibc*, 2017.
- [16] D. He, C. Westphal, and J. J. Garcia-Luna-Aceves, “Joint rate and fov adaptation in immersive video streaming,” in *Proceedings of the 2018 Morning Workshop on Virtual Reality and Augmented Reality Network*, pp. 27–32, 2018.
- [17] I. Jo, Y. Park, H. Kim, and J. Bae, “Evaluation of a wearable hand kinesthetic feedback system for virtual reality: psychophysical and user experience evaluation,” *IEEE Transactions on Human-Machine Systems*, vol. 49, no. 5, pp. 430–439, 2019.
- [18] R. Ghaznavi-Youvalari, A. Zare, H. Fang et al., “Comparison of HEVC coding schemes for tile-based viewport-adaptive streaming of omnidirectional video,” in *2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSp)*, pp. 1–6, 2017.
- [19] T. El-Ganainy and M. Hefeeda, “Streaming virtual reality content,” 2016, <http://arxiv.org/abs/1612.08350>.
- [20] Z. Tu, T. Zong, X. Xi et al., “Content adaptive tiling method based on user access preference for streaming panoramic video,” in *2018 IEEE International Conference on Consumer Electronics (ICCE)*, pp. 1–4, 2018.
- [21] M. Hosseini and V. Swaminathan, “Adaptive 360 VR video streaming based on MPEG-DASH SRD,” in *2016 IEEE International Symposium on Multimedia (ISM)*, pp. 407–408, 2016.
- [22] H. S. Kim, S. B. Nam, S. G. Choi, C. H. Kim, T. T. K. Sung, and C. B. Sohn, “HLS-based 360 VR using spatial segmented adaptive streaming,” in *2018 IEEE International Conference on Consumer Electronics (ICCE)*, pp. 1–4, 2018.
- [23] F. Guo, H. Zhang, H. Ji, X. Li, and V. C. M. Leung, “An efficient computation offloading management scheme in the densely deployed small cell networks with mobile edge computing,” *IEEE/ACM Transactions on Networking*, vol. 26, no. 6, pp. 2651–2664, 2018.
- [24] S. Henri, C. Vlachou, and P. Thiran, “Multi-armed bandit in action: optimizing performance in dynamic hybrid networks,” *IEEE/ACM Transactions on Networking*, vol. 26, no. 4, pp. 1879–1892, 2018.
- [25] K. B. Letaief, W. Chen, Y. Shi, J. Zhang, and Y. J. A. Zhang, “The roadmap to 6G: AI empowered wireless networks,” *IEEE Communications Magazine*, vol. 57, no. 8, pp. 84–90, 2019.
- [26] M. Liu and Y. Liu, “Price-based distributed offloading for mobile-edge computing with computation capacity constraints,” *IEEE Wireless Communications Letters*, vol. 7, no. 3, pp. 420–423, 2018.
- [27] Z. Liu, S. Ishihara, Y. Cui, Y. Ji, and Y. Tanaka, “JET: joint source and channel coding for error resilient virtual reality video wireless transmission,” *Signal Processing*, vol. 147, pp. 154–162, 2018.
- [28] J. Feng, Y. Wu, G. Zhai, N. Liu, and W. Zhang, “An algorithm for transmitting VR video based on adaptive modulation,” in *2019 IEEE/CIC International Conference on Communications in China (ICCC)*, pp. 443–448, 2019.
- [29] J. Chen, B. Li, and R. Srikant, “Thompson-sampling-based wireless transmission for panoramic video streaming,” in *2020 18th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOPT)*, pp. 1–3, 2020.
- [30] K. H. Lee and D. Petersen, “Optimal linear coding for vector Channels,” *IEEE Transactions on Communications*, vol. 24, no. 12, pp. 1283–1290, 1976.
- [31] B. Tan, H. Cui, J. Wu, and C. W. Chen, “An optimal resource allocation for superposition coding-based hybrid digital-analog system,” *IEEE Internet of Things Journal*, vol. 4, no. 4, pp. 945–956, 2017.
- [32] F. Q. Zhang and L. J. Ma, “Multi-object crowd real-time tracking in dynamic environment based on neural network,” *Journal of Network Intelligence*, vol. 7, no. 2, pp. 386–394, 2022.