frontiers | Frontiers in Physics

TYPE Original Research
PUBLISHED 23 July 2024
DOI 10.3389/fphy.2024.1429750

OPEN ACCESS

EDITED BY
Rajib Biswas,
Tezpur University, India

REVIEWED BY
Kemal Gokhan Nalbant,
Beykent University, Türkiye
Ivan Cvitić,
University of Zagreb, Croatia

*CORRESPONDENCE
Jaeyoung Choi,
jychoi9@gachon.ac.kr
Pi-Chung Wang,
pcwang@nch.edu.tw
Ibrahim A. Hameed,
ibib@ntnu.no

RECEIVED 08 May 2024
ACCEPTED 25 June 2024
PUBLISHED 23 July 2024

CITATION
Tiang JJ, Chung HC, Choi J, Khan I, Alshehri A, Wang P-C and Hameed IA (2024), An efficient algorithm for resource optimization in TWDM passive optical network using a C-RAN. *Front. Phys.* 12:1429750. doi: 10.3389/fphy.2024.1429750

An efficient algorithm for resource optimization in TWDM passive optical network using a C-RAN

Jun Jiat Tiang[1], Hee Chan Chung[2], Jaeyoung Choi[2]*, Imran Khan[3,4], Asma Alshehri[5], Pi-Chung Wang[6]* and Ibrahim A. Hameed[7]*

[1]Centre for Wireless Technology, Faculty of Engineering, Multimedia University, Cyberjaya, Malaysia, [2]School of Computing, Gachon University, Seongnam-si, Republic of Korea, [3]Department of Electrical Engineering, University of Engineering and Technology, Peshawar, Pakistan, [4]Islamic University Centre for Scientific Research, The Islamic University, Najaf, Iraq, [5]Department of Computer Sciences, College of Computer Engineering and Sciences, Prince Sattam Bin Abdulaziz University, Al Kharj, Saudi Arabia, [6]Department of Computer Science and Engineering, National Chung Hsing University, Taichung, Taiwan, [7]Faculty of Information Technology and Electrical Engineering, Norwegian University of Science and Technology, Trondheim, Norway

The traditional base station in C-RAN is divided into three parts: a pool of centralized baseband units (BBUs), a fronthaul network that links the BBUs and remote radio units (RRUs), and RRUs. This paper proposes a novel cooperative algorithm for resource optimization in a time-wavelength division multiplexed (TWDM) passive optical network (PON) incorporating a cloud radio access network (C-RAN). First, a joint collaborative strategy is deployed to optimize cooperative caching and transmission in the wireless and optical domains. Then, the quality of experience (QoE) is improved by bandwidth configuration and caching. Simulation results show that the average throughput of the proposed QoE-aware video cooperative caching and transmission mechanism (QACCTM) algorithm is approximately 30% higher than that of other algorithms. Compared with the relative average residual clutter power (RARCP) and quality-aware wireless edge caching (QAWEC) algorithms, the proposed QACCTM algorithm reduces the access delay by approximately 27.1% and 15.9%, respectively.

KEYWORDS

photonics, optical communication, passive optical network, cloud radio access network, optoelectronics, semiconductor

1 Introduction

With the widespread application of multimedia intelligent terminal equipment, video services are showing explosive growth trends, and the traditional distributed radio access network (D-RAN) cannot meet the ever-growing user and network requirements. To solve the above problems, researchers proposed a cloud radio access network (C-RAN) architecture based on centralized processing, cooperative radio, and real-time cloud computing [1, 2]. Different from D-RAN, C-RAN breaks the fixed connection relationship between the remote radio frequency head (RRH) (also called a remote radio unit) and baseband processing unit (BBU) so that RRH is distributed in each cell site and BBU is concentrated in the central computer room [3, 4]; a fronthaul is formed

between the BBU and the RRH [5, 6]. The fronthaul capacity is crucial to the services carried out by C-RAN. Due to the characteristics of high energy efficiency, low delay, and high transmission capacity of the time division wavelength division multiplexing passive optical network (TWDM-PON), it can meet the high bandwidth and delay-sensitive characteristics of video services, thus becoming the C-RAN optical fiber fronthaul [7–9].

Nowadays, with the development of the network, more and more applications are emerging, such as high-speed multimedia services, interactive games, and AR/VR [10–13]. These computing-intensive and delay-sensitive applications make it difficult for mobile terminals (stations), relying solely on their own computing and storage capabilities, to meet user needs. At the same time, the battery capacity of a station (STA) is relatively limited [14–17]. Therefore, offloading computing tasks to edge cloud servers that are closer to users has become a promising solution. Mobile edge computing (MEC) not only meets the expansion requirements of STA computing capabilities and improves user quality of experience (QoE) but also prolongs the service life of stations [18, 19]. Although MEC itself has many advantages, it still needs to be fully utilized under a reasonable network architecture. Among them, the fiber-wireless (FiWi) hybrid access network combines the high bandwidth and high reliability of the passive optical network (PON) with the flexibility of the wireless mesh network (WMN) [20–23], making it a promising platform for MEC implementations [24, 25]. Although the FiWi network supporting MEC offers many advantages, transferring too many computing tasks to offloading will lead to excessive communication overhead; STA needs to consume more energy to perform the uploading of offloading tasks and also needs to open more mesh nodes and optical network units (ONUs) to forward offloading data, resulting in excessive power consumption in the system [26]. The amount of data offloaded by MEC has become a key factor affecting the energy consumption of the system. In order to fundamentally control the communication overhead of offloading and reduce the energy consumption of the offloading process, addressing how to reduce the amount of offloaded data is an important problem that needs to be solved [27, 28].

Studies have shown that the number of times a user clicks on a video is closely related to the popularity of the video, and both obey the Zipf distribution [29–31]. A large number of users continuously obtain popular videos from content servers, which makes the data traffic carried by the network extremely large, resulting in a sharp decline in user QoE. Therefore, efficient wireless edge caching strategies are studied to reduce data redundancy. It is imperative to improve mobile network performance using redundant transmission [32]. For video services, access delay and wireless link rate are the main factors affecting the user quality of experience [33–35]. Therefore, optimizing the caching and transmission of video content is crucial to improving user QoE. At present, RRH edge caching is widely used to provide users with video transmission while reducing the user's access delay. [29] proposed a joint edge cache and cloud cache strategy within the C-RAN architecture, which reduces the average delay and improves user QoE. This strategy involves caching video content for users during non-traffic peak periods to reduce the bandwidth occupation pressure caused by pre-caching. However, due to the strong dynamic characteristics of users' mobile patterns and requested content, static pre-caching strategies cannot meet the time-varying needs of users. [30] proposed a joint configuration of bandwidth and buffer resources in a software-defined orthogonal frequency division multiplexing (OFDM) passive optical network. This approach involves allocating part of the bandwidth of the optical line terminal (OLT) to delay-sensitive popular videos and using the remaining bandwidth for ONU pre-buffering of the popular video in the next time slot, thereby reducing the service response delay. However, this solution does not consider the user characteristics in the ultra-dense deployment scenario [31]; in other words, users at the edge of the cell are severely interfered with by the same layer, and the unbalanced cell load will lead to poor user performance in the heavily loaded cell. Aiming at the intensive cell deployment scenario, [32] proposed using the existing cache of the base station to perform joint transmission clustering for edge users to minimize the amount of returned data. However, considering the joint transmission under the given buffer resources cannot guarantee the QoE for edge users, and the cache utilization is low. [33] determined the resolution version of the video stream by considering the quality difference between video segments and the information on the user's buffer to ensure the best viewing quality for a single user without interrupting the playback, but this only optimizes a single video. When multiple users compete for resources over a shared link, unfair wireless resource allocation occurs.

In view of the above shortcomings, under the joint architecture based on TWDM-PON and C-RAN, this paper proposes a QoE-aware video collaborative caching and transmission mechanism.

The main contributions are as follows:

- Using the inter-cell interference degree, request content similarity, and ONU transmission bandwidth to represent optical and wireless domain cooperative buffering and transmission gain, a virtual passive optical network (VPON) is constructed through RRH and ONU clustering.
- Implementation of joint optimization of video stream QoE, bandwidth configuration, and cache allocation: in the optical domain, the dynamically pre-cache video content is analyzed and the cache value and dispersion of video are calculated to reduce access delay.
- In the wireless domain, the wireless bandwidth is flexibly configured based on the buffer level of each user to improve the quality of the video users watch.

The remainder of this paper is organized as follows: in Section 2, the proposed system model is discussed in detail. In Section 3, the VPON construction based on the cooperation gain approach is described. In Section 4, the joint fronthaul bandwidth and buffer allocation mechanism is discussed. In Section 5, the QoE-aware wireless bandwidth allocation is presented. In Section 6, the simulation results are discussed. Section 7 provides the conclusion of the paper.

2 Collaboration architecture of the optical and infinity domains

The joint architecture based on TWDM-PON and C-RAN proposed in this paper is shown in Figure 1. The optical network

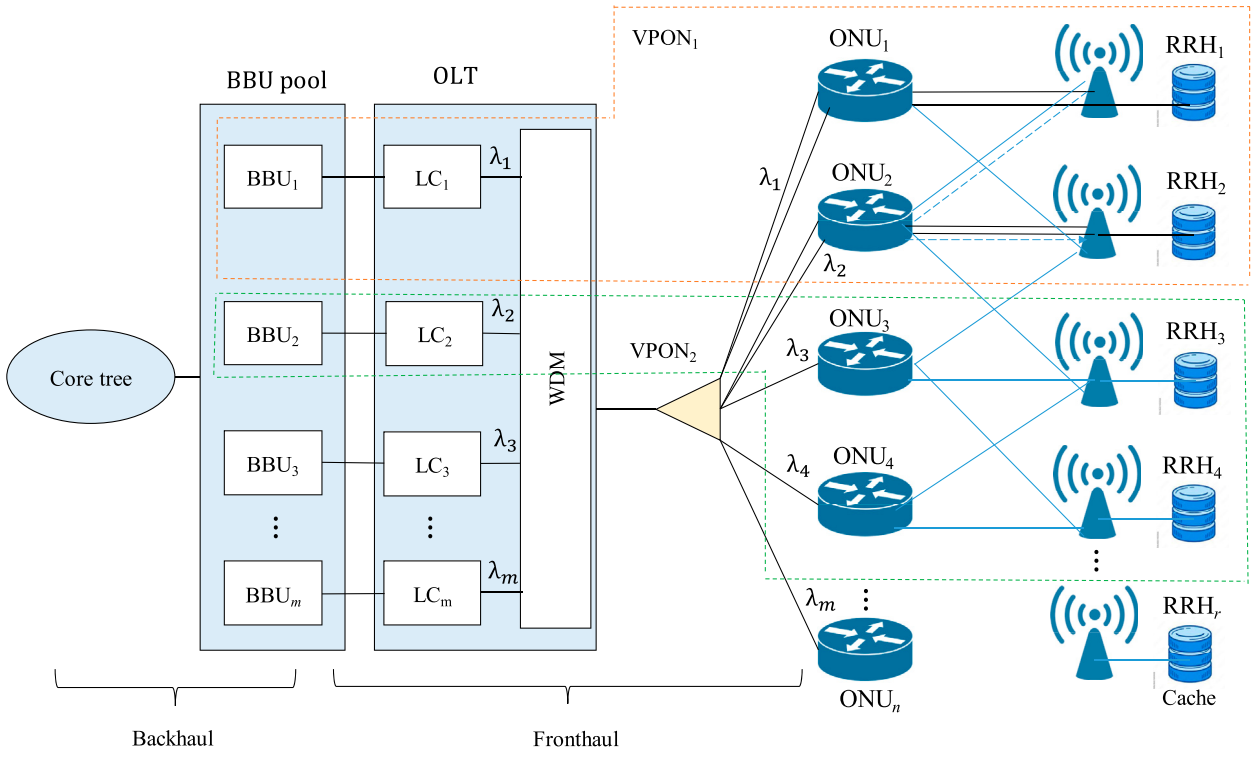

FIGURE 1
Proposed system model.

adopts a point-to-multipoint tree structure and is composed of OLT and multiple ONUs. The wireless network is composed of a gateway node (ONU-RRH, optical network unit-remote radio head) that integrates ONU and radio frequency functions.

Different from other existing architectures, each ONU in the proposed architecture can connect multiple RRHs. The connection between the BBU pool and the ONU-RRH constitutes the fronthaul, and the connection with the core network constitutes the backhaul. Based on the strong transmission capacity of TWDM-PON, multiple ONUs can use their wavelength tuners to tune to the same wavelength for wavelength sharing [36–38]. At the optical line terminal, the access to the wavelength is controlled by the line card (LC). The combination of LC and multiple ONU-RRHs communicating on its control wavelength is called a virtual PON. As shown in Figure 1, the short and long dashed boxes form $VPON_1$ and $VPON_2$, respectively. It can be seen that using the wavelength division characteristics of TWDM-PON can establish multiple VPONs and perform optical coupling and separation through wavelength division multiplexers [39].

To improve the QoE of users in the cell edge and heavy-load cells, this paper proposes the cooperative buffering and transmission of the optical and wireless domains, which is mainly realized by constructing an optical serving cell (OSC) and a wireless serving cell (WSC) [40, 41]. The wireless domain serving cell is a collection of RRHs centered on edge users, which is used to improve the wireless link rate of edge users. Benefiting from the architecture of VPON, RRH in WSC provides services for users through joint transmission (JT). When operating in the JT mode, several RRHs transmit data for a single user on the same time–frequency resource block and transform inter-cell interference signals into beneficial signals to enhance the performance of cell edge users [32]. Taking advantage of the joint transmission, the video requested by the cell edge user can be distributed and cooperatively cached in the serving cell of the wireless domain. However, when the load of the RRH is large, due to the limited buffer space, many users still need to obtain videos through fronthaul and backhaul, which is prone to congestion and reduces user QoE [42, 43]. Therefore, the optical domain serving cell is represented as the RRH of user cooperative caching, which is used to improve the content access delay of the users in the heavy-loaded cell and prevent a large number of users in the heavy-loaded cell from accessing the content server to obtain videos. As shown in Figure 1, there are two transmission modes in the optical domain service cell: one is the cooperative transmission of caching through the forwarding mode of RRH–ONU–BBU–ONU–RRH [34], as shown by the thick arrow in Figure 1 [44, 45]. Since the ONU and RRH of the architecture proposed in this paper have a one-to-many connection relationship, the other can use ONU as a forwarding medium to perform cached cooperative transmission through the path of RRH–ONU–RRH, as shown by the dotted arrow in Figure 1.

3 VPON construction based on cooperation gain

In a dense cell deployment scenario, users at the edge of the cell are covered by multiple RRHs and experience severe

co-channel interference, and the unbalanced load among cells affects the cache utilization of lightly loaded RRHs. However, a large number of users of heavy-loaded RRH still need to obtain videos through the core network, and the access delay is large. This section proposes a VPON construction mechanism of cooperative buffering and transmission gain awareness. The construction of VPON is mainly realized through RRH and ONU clustering. Among them, RRH clustering realizes wireless domain cooperative buffering and transmission gain, and ONU clustering realizes optical domain cooperative buffering and transmission gain.

3.1 RRH clustering

In the VPON architecture proposed in this paper, joint transmission within VPON or VPONs can be used to realize joint transmission of RRHs. The former only needs to interact between cells in the same BBU to process related business data and control information, making implementation easier. The latter needs exchanging this information between BBUs, which places high requirements on the bandwidth of the X2 interface and results in a greater delay than the former [35]. Therefore, this paper only considers joint transmission within VPON. Before a user establishes a wireless domain serving cell, it is necessary to cluster the RRHs to maximize the wireless domain coordination cache and transmission gain of the edge users in the cooperation area [46, 47]. Based on the breadth-first search (BFS) method, this paper proposes a greedy dynamic clustering algorithm. First, the BBU builds a connection relationship graph according to the similarity between the interference information on the edge users and the request content of the cell. Second, the RRH is clustered based on the edge weight, that is, the connection relation graph is divided into multiple subgraphs, and each subgraph represents a joint transmission cluster.

3.1.1 Graph construction based on edge weights

The received signal y_k at the time of service can be obtained using Eq. 1.

$$y_k = h_{i,k}P_k + \sum_{j\neq i} h_{j,k}P_j + z_k, \tag{1}$$

where $h_{i,k}$ and $h_{j,k}$ represent the channel gain from RRH_i and RRH_j to user k, respectively; P_k and P_j represent the useful and interference signal flows, respectively; and z_k represents the Gaussian white noise of user k. User k is determined using the local RRH_i

Let I_{r_i,r_j} denote the degree of interference between RRH_i and RRH_j, that is, the average value of interference signals received by marginal users in RRH_i from RRH_j. The calculation expression is shown in Eq. 2.

$$I_{r_i,r_j} = \frac{\sum_{k\in u_i^e} \mathrm{RSRP}_{j,k}}{\left|u_i^e\right|}, \tag{2}$$

where $\mathrm{RSRP}_{j,\ k}$ represents the received power of the reference interference signal received by user k from RRH_j and $|u_i^e|$ represents the number of marginal users in RRH_i who receive the interference signal from RRH_j.

The similarity S_{r_i,r_j} of the edge user request content between RRH_i and RRH_j can be obtained using the Jaccard coefficient, as shown in Eq. 3.

$$S_{r_i,r_j} = \frac{\left|N(r_i) \bigcap N\left(r_j\right)\right|}{\left|N(r_i) \bigcup N\left(r_j\right)\right|}. \tag{3}$$

If we use v_k^i to represent the binary variable of video k requested by edge users in RRH_i, then $N(r_i) = \{v_1^i, v_2^i, \ldots, v_k^i, \ldots, v_K^i\}$ is the video collection requested by RRH_i, where K is the total number of videos. It can be seen from Eq. 3 that $0 \le S_{r_i,r_j} \le 1$; the larger the value of S_{r_i,r_j}, the more similar the interest preferences of the two cell edge users.

Therefore, the cooperative caching and transmission gain w_{r_i,r_j} from $\mathrm{RRH_i}$ to $\mathrm{RRH_j}$ can be expressed using Eq. 4.

$$w_{r_i,r_j} = I_{r_i,r_j}S_{r_i,r_j}, \tag{4}$$

where the interference degree I_{r_i,r_j} represents the cooperative transmission gain between two cells and the request content similarity S_{r_i,r_j} represents the cooperative caching gain.

The graph $G = (V, E)$ is constructed according to cooperative caching and transmission gain. RRH represents the vertex, and its weight is the load of the RRH (load_v). The weight of the edge between two RRHs is the cooperative caching and transmission gain, which is expressed as

$$w_{r_i,r_j} = w_{r_i\to r_j} + w_{r_j\to r_i}. \tag{5}$$

When $w_{r_i,r_j} = 0$, no connection needs to be built. Each RRH has a list of weights with other RRHs, $W_i = \{w_{i,1}, w_{i,2}, \ldots, w_{i,R}\}$. The sum of directed edge weights between two vertices in graph G indicates that two RRHs are included in the same cluster of collaborative caching and transmission gain. So, the directed graph is converted into an undirected graph using Eq. 5, where $w_{r_i\to r_j}$ and $w_{r_j\to r_i}$ are the weights of the directed graph.

3.1.2 Graph-based RRH clustering

When RRH is clustered, for any two vertices i and j, if the condition shown in Eq. 6 is met—where i and j are the maximum edge weights of their adjacent vertices—then the two vertices are placed in the same cluster to achieve maximum cooperative cache and transmission gain within the cluster.

$$\begin{cases} w_{r_i,r_j} = \max\limits_{r_s} w_{r_i,r_s} \\ w_{r_i,r_j} = \max\limits_{r_k} w_{r_k,r_j} \end{cases}. \tag{6}$$

Here, s and k represent the adjacent vertices of i and j, respectively. Similarly, this rule can be extended to points to subgraphs and subgraphs to subgraphs, and then, the weights of points to subgraphs and subgraphs to subgraphs are calculated using Eq. 7.

$$w_{i,G_1} = \sum_{k\in G_1} w_{i,k}, w_{G_1,G_2} = \sum_{i\in G_1}\sum_{j\in G_2} w_{i,j}. \tag{7}$$

This paper proposes a greedy clustering algorithm based on breadth-first search, as shown in Algorithm 1. First, each vertex is regarded as a subgraph of a single vertex [48, 49]. In each search phase, we started from each subgraph, visited the subgraph (that

is, find its adjacent subgraph), and selected the edge weight with the maximum adjacent subgraphs that are used as their candidate collaboration points. If the two subgraphs satisfy Eq. 6, that is, the adjacent subgraph of the maximum edge weight of the candidate cooperation point is the subgraph itself, then the candidate cooperation point and the subgraph are placed in the same cluster, and the distance between the two subgraphs is removed. Then, we continued to visit the candidate collaboration points in the new subgraph [50, 51]. This procedure is repeated until the submap exceeds the wavelength capacity. The algorithm terminates when all subgraphs are included in the cluster (i.e. $S = \varnothing$). According to Algorithm 1, the graph will be divided into multiple subgraphs; each subgraph represents a joint transmission cluster, and the maximum cooperative caching and transmission gain can be achieved within each cluster.

```
Treat each vertex v ∈ V as a subgraph, and all subgraphs form a set S = {v}. The set of RRH clusters is VP, and VP = ∅ is the search phase.
While (V ≠ ∅ and W ≠ ∅)
For each subgraph G ∈ S, search its adjacent vertex k ∈ V and adjacent subgraph G_a ∈ S, and find out the vertex or subgraph Ω(G) with the largest weight with this subgraph.
For each Ω(G)
if G = Ω(Ω (G) and load_{G ⋃ Ω(G)}) ≤ C_W
G = G ⋃ Ω(G)
Remove Ω(G) from the set S
end if
if G = Ω(Ω (G)and load_{G ⋃ Ω (G)}) > C_W
  Remove G from the S set and add it to VP
  end if
  Update the edge weights between connected subgraphs according to Eq. 7
  end for
  Remove the remaining subgraphs in S to the cluster set VP
  end while
```

Algorithm 1. RRH greedy clustering.

3.2 ONU clustering

For video services, playback interruption is a key factor affecting user QoE. Constrained by the limitation to RRH radio resources, during the maximum initial buffering time of the initial buffering stage, when the actual buffered data on the user are less than the expected buffered data (i.e., $B_t < B_e$), video interruption occurs. In this case, in order to prevent interruptions of video playback, users need to switch the resolution of the video. Clearly, there is a tight relationship between video resolution, maximum initial buffer time, and network load [52, 53]. At present, the rate-adaptive algorithm based on the E-LBP table is generally used to define the load situation of the RRH, that is, to characterize the load situation of the RRH by whether the user in the cell switches the video bit rate version [54–56]. Different video resolutions and transmission rates in the E-LBP table correspond to different maximum initial buffering times $D_{\max}$.

The light and heavy load classification of RRHs is shown in Eq. 8. If there is a user switching bit rate version in RRH_r, it indicates that the RRH is overloaded and users have to switch resolution versions to avoid interruption. When all users in the RRH do not change the resolution version within the maximum initial buffer time, it is a lightly loaded RRH.

$$\begin{cases} r \in VP_m^h, \exists\, u \in U_r, c(u) = 1 \\ r \in VP_m^l, \forall\, u \in U_r, c(u) = 0 \end{cases}, \tag{8}$$

where $c(u)$ is defined as the binary variable of whether user u switches the resolution version and VP_m^h represents the cluster set.

First, the minimal effective dominating set of each VPON should be solved, and all effective dominating sets can be calculated using Eq. 9. Then, the minimal effective dominating set of each VPON should be solved, and all effective dominating sets can be calculated using Eq. 9.

$$H = \left(v_1^1 \vee v_2^1 \vee \ldots v_1^1\right) \wedge \left(v_1^2 \vee v_2^2 \vee \ldots v_{t_2}^2\right) \wedge \ldots \wedge \left(v_1^n \vee v_2^n \vee \ldots v_{t_n}^n\right). \tag{9}$$

ONU and RRH have a one-to-many connection relationship, and RRHs in different VPONs may be connected to the same ONU. However, ONUs can only be tuned to one wavelength in the same scheduling cycle, so it is necessary to cluster the ONUs so that each VPON can run independently. The clustering of ONUs must make each RRH dominated by at least one ONU [57–59]; then, the clustering of ONUs can be transformed into the dominating set problem of the bipartite graph. All RRH nodes in any VPON $V = \{v_1, 2_2, \ldots, v_m\}$ form a set $S = \{s_1, s_2, \ldots, s_n\}$. The links among them form a set E, and then, the relationship between ONU and RRH can be represented using a bipartite graph $G = (S, E, V)$. The RRH node s_i in any VPON is dominated by the connected ONU node $v_1^i, v_2^i, \ldots, v_{t_i}^i \in V$ number of ONUs.

Equation 9 indicates that the combination formed by selecting any ONU node v_j that dominates RRH node s_i is an effective dominating set. Furthermore, Eq. 9 is simplified according to the absorption law of the dominating set to solve the minimal effective dominating set.

When the RRHs connected to the ONU are in different VPONs, the weight Ω_{mn} of the ONU in different VPONs is expressed in Eq. 10:

$$\Omega_{mn} = \alpha_m \frac{\sum_{r_i \in R_n^h} \sum_{r_j \in P_m^l} \frac{S_{r_i r_j}}{\left|VP_m^l\right|\left|VR_n^h\right|}}{M_n \mu_m}. \tag{10}$$

All VPONs need to allocate the remaining ONU nodes after finding the minimum effective dominating set. Due to the limited cache space, the light-loaded RRH can provide optical domain cooperative cache space for heavy-loaded RRH users to enhance the cache hit rate [60–62]. The maximum transmission capacity of ONU_n is defined as M_n, and μ_m indicates that $\mathrm{ONU_n}$ only connects to the bandwidth utilization rate of RRH in VPON_m. The remaining bandwidth of the ONU is used to represent the optical domain cooperative transmission gain, and the average content request similarity of light and heavy load RRH is used to represent the optical domain cooperative cache gain. R_n^h represents the set of

heavily loaded RRHs connected to ONU_n in $VPON_m$; V denotes the set of VPONs [63]; P_m^l denotes the signal power of $VPON_m$; α_m represents the average load rate of all ONUs in $VPON_m$; and S_{r_i,r_j} is the similarity of the edge user request content between RRH_i and RRH_j. In the case of the same optical domain cooperation gain, the ONU is allocated to the VPON with a larger average load rate to achieve ONU load balancing. The gain of ONU_n in $VPON_m$ is inversely proportional to the amount of bandwidth used by the ONU [64, 65]. The higher the ONU load, the greater the queuing delay and the smaller the optical domain cooperative transmission gain. The gain of ONU_n in $VPON_m$ is proportional to the average content request similarity between light and heavy load RRHs. The ONU is included in the VPON with the maximum gain, and then, the inherent connection between the ONU and RRH in other VPONs is disconnected by closing the interface of the ONU, thereby forming an independent VPON [66].

4 Joint fronthaul bandwidth and buffer allocation mechanism

The QoE refers to the user's comprehensive subjective experience of the quality and performance of applications or services, reflecting the user's acceptance of services. The initial access delay directly affects the user's satisfaction with the video. If the access delay is too long, the user may give up playing the video. The static cache configuration method cannot meet the time-varying needs of users. This paper proposes a joint fronthaul bandwidth and cache allocation mechanism to reduce the user access delay in the next scheduling cycle [67]. To this end, first, the cache value of the video is calculated by considering user behavior and collaborative caching of the light domain. Second, according to the connection state of ONU-RRH, the dispersion of the requested video is defined, which is used to evaluate the bandwidth utilization of the pre-caching process. Finally, pre-caching of videos is accomplished through multiple iterations based on the value-to-cost ratio of video caching.

4.1 Video caching value analysis

Since most video requests focus on a small number of video content, user video requests obey the Zipf distribution with parameter $\tau = 0.56$. The probability of user requesting video v is expressed in Eq. 11

$$f_v = \frac{1/\tau}{\sum_{m=1}^{V} 1/m^{\tau}}. \tag{11}$$

Here, $\tau = 0$ indicates that the video popularity follows a uniform distribution and the larger τ values indicate that video requests are concentrated on fewer popular videos.

The value of the video cache is closely related to user behavior, so χ_{ir} is defined as the probability that the cached content in RRH_r is hit by user i. This parameter depends on the user i's movement pattern (i.e., current location, movement speed, and movement direction) and the request video's popularity [68, 69]. When a given user requests the popularity of a video, the faster he moves or is at the edge of the cell, the smaller the probability of being served by RRH_r. When the user's mobility pattern is given, the higher the popularity of the requested video, the higher the probability of using that RRH to cache the video. χ_{ir} can predict user behavior through machine learning, so the probability set $\{\chi_{ir}\}$ of user request cache is known to the system. According to [70], the echo state network model can be used to predict the probability p_{iv} of user i requesting video v in the scheduling period and the probability s_{ir} of user i being served by RRH_r [71–73], so the probability that the cache content in RRH_r is hit by user i can be calculated using Eq. 12, where $\chi_{ir} \in (0, 1)$.

$$\chi_{ir} = f_v p_{iv} s_{ir}. \tag{12}$$

Light-loaded RRHs can provide optical domain cooperative buffering for heavy-loaded RRHs. To avoid heavy-loaded RRHs being overloaded, the buffer bandwidth for heavy-loaded RRH_r to receive other light-loaded RRHs is defined as B_r. Assuming that the size of video v is s_v, the required transmission bandwidth in the scheduling period is b_v. The buffer space of RRH_r satisfies the following two conditions, and video v needs to be pre-cached.

1) Direct caching conditions: the size of the remaining space in the cache is $C_r^{'} > 0$ and $C_r \geq s_i$.
2) Cache replacement condition: condition 1) is not satisfied, but the request probability of cached video v' is smaller than that of video v, and $s_{v'} > s_v$ is satisfied. Otherwise, the sum of the request probabilities of n videos is smaller than the request probability of video v and satisfies $s_1 + \ldots + s_n > s_v$.

For heavy-load RRH_r, if the video v satisfies the above conditions, it will be cached, and its cache value is defined as g_{rv}^h. The calculation expression is expressed in Eq. 13

$$g_{rv}^h = \frac{g_{rv}^1}{s_v} = \frac{\sum_{i \in I_r} a_{iv} \chi_{ir}}{s_v}, \tag{13}$$

where a_{iv} is a binary variable, indicating whether user i requests video v; s_v is the cache space of video v; and I_r is the user set in RRH_r. The cache value is proportional to the probability of a user requesting a cached video v in RRH_r and g_{rv}^1 and is inversely proportional to the cache space s_v occupied by the video v. χ_{ir} is the probability that the cached content in RRH_r is hit by user i, which can be determined using Eq. 12.

In VPON, it is assumed that the numbers of heavy- and light-load RRHs are M_1 and M_2, respectively [71, 72]. For the heavily loaded RRHs with the possibility of optical domain cooperative caching in VPON, the sum of the probabilities of requesting caching video v is gc_v, which is expressed in Eq. 14

$$gc_v = \sum_{r=1}^{M_1} (1 - b_{rv}) \sum_{i \in I_r} a_{iv} \chi_{ir} \left[B_r^{'} \geq b_v \right], \tag{14}$$

where b_{rv} is a binary variable, indicating whether video v satisfies the condition of buffering in RRH_r. When the remaining receiving bandwidth $B_r^{'}$ of heavy-load RRH_r is greater than the transmission bandwidth of video v, $[B_r^{'} \geq b_v] = 1$; otherwise, $[B_r^{'} \geq b_v] = 0$. For heavy-load RRHs with optical domain cooperative caching, video v selects light-loaded RRH_r with a maximum buffering value for cooperative caching. Then, the optical domain serving cell of the user requesting video v in heavy-loaded RRH is RRH_r [73]. Thus, the cache value of light load RRH_r can be calculated using Eq. 15.

$$g_{rv}^{l} = \frac{g_{rv}^{1} + gc_{v}}{s_{v}}, \tag{15}$$

where s_v is the cache space occupied by the video v and gc_v is defined in Eq. 14. According to the above analysis, the cache value of video v in $VPON_m$ can be calculated using Eq. 16.

$$gp_{v} = \sum_{r \in VP_{m}^{h}} b_{rv} g_{rv}^{h} + \sum_{r \in VP_{m}^{l}} b_{rv} g_{rv}^{l}, \tag{16}$$

where VP_m^h is the cluster of $VPON_m$ and b_{rv} is defined in Eq. 14.

4.2 Request video discreteness

Optical domain ONUs and wireless-domain RRHs have a one-to-many connection relationship. Similarly, RRHs can also be connected to multiple ONUs. For multicast video services, if multiple RRHs apply to cache the same video at the same time, a corresponding multicast routing table will be generated in the network, and different video services correspond to different multicast routing tables. The more dispersed the RRH of the requested video v is, the greater the need for ONU dominance points and the easier it is to cause redundant transmission of data and waste the bandwidth resources of the ONU. Therefore, this paper defines the dispersion degree MD_v of the requested video according to the connection state of the ONU–RRH, which reflects the utilization rate of the ONU transmission bandwidth when the video is cached, and the calculation expression is shown in Eq. 17.

$$MD_{v} = \frac{N_{v}}{N_{m}}, \tag{17}$$

where N_v represents the minimum number of ONUs required to dominate the RRH requesting to buffer video v; N_m represents the total number of ONUs in $VPON_m$; and $0 < MD_v \leq 1$. The smaller the dispersion of video v, the higher the concentration of RRHs requesting video v, leading to better multicast performance. Multicast through ONUs can effectively improve bandwidth utilization and reduce redundant data transmission.

4.3 Joint fronthaul bandwidth and the buffer allocation algorithm

The dynamic pre-caching process is not only related to the cache value of video but also considers the bandwidth cost of pre-caching, including the downlink wavelength bandwidth and ONU transmission bandwidth. Therefore, the value-to-cost ratio ϕ_v of cached video v is defined, as shown in Eq. 18.

$$\phi_{v} = \frac{gp_{v}}{MD_{v} b_{v}}, \tag{18}$$

where b_v is the required transmission bandwidth in the scheduling period; gp_v is the cache value calculated from Eq. 16; and MD_v is calculated using Eq. 17. The value-to-cost ratio is directly proportional to the cache value of the video and inversely proportional to the transmission bandwidth required by the video and the discreteness of the video.

The cache value in Section 4.1 does not consider the availability of fronthaul bandwidth and ONU transmission bandwidth. In other words, when the fronthaul bandwidth or the remaining bandwidth of the ONU is less than the transmission bandwidth of video v, RRHr does not have the value of caching video v,. Therefore, Eq. 16 can be rewritten as

$$\begin{aligned} gp_{v} &= \sum_{r \in VP_{m}^{h}} b_{rv} g_{rv}^{h} + \sum_{r \in VP_{m}^{l}} b_{rv} g_{rv}^{l}, \\ \text{s.t.}\, C_{w}^{'} &\geq b_{v}, \exists\, c_{or} B_{o}^{'} \geq b_{v}, \end{aligned} \tag{19}$$

where $C_w^{'}$ represents the remaining wavelength capacity, $B_o^{'}$ represents the remaining transmission bandwidth of ONU_o, and c_{or} represents the binary variable of the connection status between ONU_o and RRH_r. According to Eq. 19, the value-to-cost ratio of video v can be calculated, and the joint fronthaul bandwidth and cache allocation algorithm can pre-cache according to the descending order of the value-to-cost ratio of each video. The algorithm preferentially selects the video v with the highest value-to-cost ratio for caching, and the minimum dominant number N_v must satisfy the ONU's transmission bandwidth constraints. Second, gc_v, g_{rv}, and MD_v are updated, respectively, according to the remaining receiving bandwidth $B_r^{'}$, remaining fronthaul bandwidth $C_w^{'}$, and remaining ONU transmission bandwidth $B_o^{'}$ of the heavily loaded RRH, and ϕ_v is recalculated. Finally, the video is precached for multiple iterations until the cache does not need to be updated or the bandwidth is exhausted.

5 QoE-aware wireless bandwidth allocation

The allocation of wireless bandwidth not only affects the user's initial access delay but also directly determines the user's perception of video quality. Therefore, this section proposes a QoE-aware wireless bandwidth allocation mechanism. According to the buffer status information obtained in Section 4, first, a wireless domain serving cell is established for edge users to associate RRH with users, thereby reducing user access delay and ensuring a high signal-to-noise ratio. Second, the QoE optimization of each user is independent and selfish, which leads to low utilization of wireless resources and poor QoE. Therefore, this paper proposes buffer-level-aware wireless bandwidth allocation to realize flexible configuration of wireless resources among users and further improve QoE.

5.1 Establishment of the user wireless domain serving cell

To improve the user's QoE, the wireless domain serving cell provides two transmission modes for edge users in combination with the buffer state of the RRH. As shown in Figure 2A, when all RRHs in the serving cell of the wireless domain cache the content requested by the user, all RRHs provide video services for the user through joint transmission. As shown in Figure 2B, when some RRHs in the serving cell of the wireless domain have cached the video requested by the user, the cached part of the RRHs is used for joint transmission for the user, and the RRHs that have not cached the video are prevented from using the video resource block to

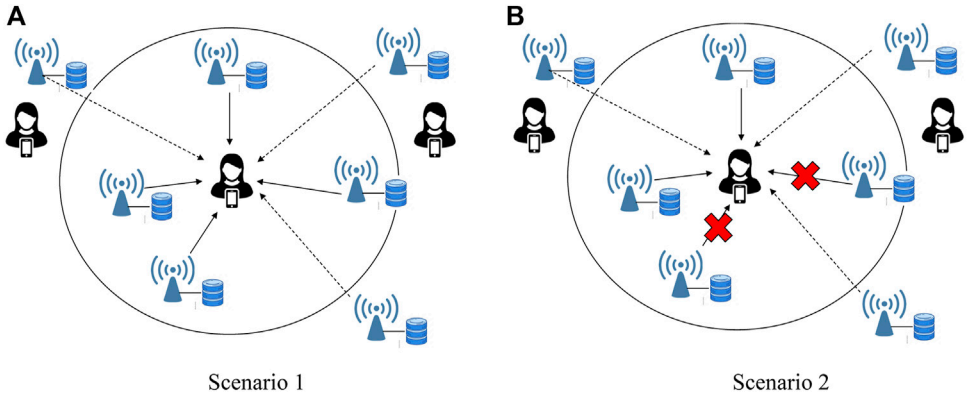

FIGURE 2
System model of the wireless domain serving cell cooperative transmission mode: **(A)** Scenario 1. **(B)** Scenario 2.

eliminate interference. This method not only improves the user's wireless data rate but also avoids the delay caused by the uncached RRH to the content server to obtain the video and reduces the user's access delay.

According to the two transmission modes of the serving cell in the wireless domain, the association between the edge user and the RRH can be established for the edge user through the average signal-to-interference and noise ratio $\overline{\mathrm{SIN\,R}_r}$ of the cell. Assume that the reference signal received power set of edge user i is RSRP_r, and the values are arranged in descending order. The wireless domain serving cell WSC_i for user i is established by sequentially including the corresponding RRHs in RSRP_i $\mathrm{SIN\,R}_i \geq \overline{\mathrm{SIN\,R}_r}$, and the calculation of its signal-to-noise ratio (SINR) is given in Eq. 20.

$$\mathrm{SIN\,R}_i = \frac{\sum_{r \in \mathrm{WSC}_i} c_{ir} g_{ir} p_r}{\sigma_0 + \sum_{r \notin \mathrm{WSC}_i} g_{im} p_m}, \tag{20}$$

where σ_0 represents the noise power of Gaussian white noise; g_{ir} and g_{im} represent the channel gain; p_r and p_m represent the transmit power of RRH_r and RRH_m, respectively; and WSC_i is the wireless serving cell i. By establishing wireless domain serving cells for edge users, the association set U_r^{wir} between each RRH and users is determined. When the video requested by user i has been cached in RRH_r, $c_{ir} = 1$; otherwise, $c_{ir} = 0$.

The allocation of RRH wireless bandwidth directly determines the associated user's perception of video quality. This paper uses the mean opinion score (MOS) to represent the user satisfaction of real-time video streaming. The utility function Q of the video stream is defined as a function of the sending data rate γ, as shown in Eq. 21.

$$Q = f(\gamma), f: \gamma \rightarrow \mathrm{MOS}. \tag{21}$$

According to the test report in [71], the peak signal-to-noise ratio (PSNR) and MOS can be regarded as a simple linear mapping relationship. Therefore, the MOS value can take any value between 1.0 (that is, PSNR is 30 dB) and 4.5 (that is, PSNR is 42 dB), where 1.0 represents the worst user satisfaction and 4.5 represents the best user satisfaction. Since different video sequences have different image characteristics, the utility function Q adopts the parameter model of [72], uses PSNR to measure video quality, and uses limited empirical data (i.e., three pairs of peak signal-to-noise ratio and transmission data rate value) The parameter values in the model are derived to determine the functional relationship between PSNR and γ.

Assuming that the user receives all data without errors, the wireless data rate for any user $i \in U_r^{wir}$ is shown in Eq. 22.

$$\gamma_i = \mathrm{g}(\delta_i) = \delta_i B_{\max}(1 + \mathrm{SIN\,R}_i), \tag{22}$$

where $B_{\max}$ is the maximum available bandwidth of the RRH. The goal of this paper is to determine the wireless resource share δ_i of each user in the optimization period, that is, the proportion of physical resource blocks (RBs).

The goal of a QoE-aware radio resource allocation strategy is to determine the transmission rate of users associated with RRHs, thereby maximizing user satisfaction [73]. Therefore, the optimization problem is shown in Eq. 23, which can be solved using the greedy search algorithm.

$$\begin{aligned} \delta_{\mathrm{opt}} &= \arg \max_{\{\delta_1, \delta_2, \ldots, \delta_N\}} \frac{1}{N} \sum_{i=1}^{N} \{Q_i(\mathrm{g}(\delta_i)) - P_i\}, \\ &\text{s.t.} \sum_{i=1}^{N} \delta_i = 1, \mathrm{g}(\delta_i) \geq \gamma_{\min,i}, \end{aligned} \tag{23}$$

where δ_{opt} represents the optimal resource allocation group; N is the number of users associated with each RRH ($N = U_r^{wir}$). $\mathrm{g}(\delta_i) \geq R_{\min,i}$ indicates that the allocation of wireless bandwidth must meet the most basic video quality of the user. If $\mathrm{g}(\delta_i) < R_{\min,i}$, the user cannot decode the lowest resolution version, resulting in a waste of wireless resources. P is the penalty item of user i, which represents user i's perception of quality fluctuation, as shown in Eq. 24. The large quality fluctuation of the user during the continuous scheduling period will cause the video to be played unevenly and affect the user's QoE. If the quality change between two consecutive optimization periods t and $t-1$ exceeds the just noticeable difference (JND) threshold λ_{th}, a quality fluctuation penalty term P_i will be added.

According to the test results of [71], the threshold $\lambda_{\mathrm{th}} = 0.23$ MOS is taken. When $|Q_i(\mathrm{g}(\delta_i))_t - Q_i(\mathrm{g}(\delta_i))_{t-1}| \leq \lambda_{\mathrm{th}}$, the fluctuation of the quality is within the range that the user cannot perceive.

$$P_i = \max\left\{0, \left|Q_i\left(g\left(\delta_i\right)\right)_t - Q_i\left(g\left(\delta_i\right)\right)_{t-1}\right| - \lambda_{\text{th}}\right\}. \tag{24}$$

According to the wireless rate allocation in Eq. 23, the playback resolution version φ_i is selected for the user so that $\varphi_i < \gamma_i$ to ensure uninterrupted video playback. In this mode, the user's QoE is not optimal because the buffer of the user equipment has a certain cache, and a higher resolution version can be selected according to its buffer level (BL). The buffering level is defined as the length of the video buffered in the buffer of the user end so that the range of resolution versions that the user can choose can be determined, as given in Eq. 25.

$$\gamma_{\min,i} \le \varphi_i \le \gamma_i\left(1 + \frac{\text{BL}_i}{T}\right). \tag{25}$$

Equation 25 indicates that during the optimization period T, in order to ensure uninterrupted video playback, the amount of input data must be greater than or equal to the amount of data consumed by the user, that is, $\varphi_i T \ge \gamma_i(\text{BL}_i + T)$. Therefore, the highest resolution version of user i can be determined according to the resource allocation γ_i obtained from Eq. 23 and the user's buffer level BL_i so as to maximize user satisfaction.

$$\begin{aligned} \varphi_i^{\max} &= \arg\max_{\varphi_i} Q_i[\varphi(\gamma_i, \text{BL}_i)], \\ \text{s.t. } &\gamma_{\min,i} \le \varphi_i \le \gamma_i\left(1 + \frac{\text{BL}_i}{T}\right). \end{aligned} \tag{26}$$

Buffer level-aware resolution version selection is conducive to improving user satisfaction. Although selecting the highest bit rate version according to Eq. 26 increases user satisfaction, the dynamic changes in the wireless channel are not conducive to users with lower buffer levels. It can be seen that the construction of the buffer level can compensate for the dynamic changes in the wireless channel. Achievable QoE indicates the MOS value that the user can achieve without considering the buffer level. Users who can achieve higher QoE but have lower buffer levels can switch to the lower bitrate version, filling the buffer. Conversely, users who can achieve lower QoE but have higher buffer levels can switch to a higher resolution rate version. Therefore, this paper defines the upper and lower limits of QoE associated with the buffer level, which are represented by UB and LB, respectively. In order to simplify the complexity of the problem, UB, LB, and the buffer level BL_i are linearly proportional. It is expressed in Eq. 27:

$$\begin{cases} \text{LB} = Q_L + a\,\text{BL}_i \\ \text{UB} = Q_H + b\,\text{BL}_i \end{cases}, \tag{27}$$

where Q_L and Q_H take the values of 1 and 4, respectively; a and b represent the user's buffer consumption speed and buffer creation speed, respectively. When $Q_L = 1$, the corresponding LB indicates the minimum MOS value that the user can achieve with the increase in the buffer level at the lowest transmission rate; $Q_H = 4$ guarantees that the quality of users with favorable radio resources but low buffering levels will not be degraded too much. According to the transmission rate γ_i in Eq. 23, the MOS value achievable by the user is $\mu_i = Q_i(\varphi_i)$, where $\varphi_i \le \gamma_i$. Therefore, the buffer level threshold κ_{th} is defined according to the value range of μ_i.

1) When $Q_L \le \mu_i \le Q_H$, that is, when the user is in a low QoE, the buffer level threshold κ_{th} is calculated as shown in Eq. 28, where a determines the speed at which user i consumes the buffer.

$$\kappa_{\text{th}} = \frac{\mu_i - Q_L}{a}. \tag{28}$$

Equation 28 shows that the higher the user's transmission rate or achievable MOS value, the greater the buffer level threshold, the faster the buffer consumption, and the smaller the buffer level threshold. Therefore, the user-selected resolution version can be regulated by the buffer threshold, as shown in Eq. 29. When the buffer level is greater than the threshold, it indicates that there is enough buffer to support the user to watch the higher bit rate version φ_i. When the buffer level is below the threshold, the lowest-resolution version is kept above the achievable transfer rate.

$$\begin{cases} Q_i(\varphi_i) \ge Q_L + a\,\text{BL}_i, \text{BL}_i \ge \kappa_{\text{th}} \\ Q_i(\varphi_i) \ge \mu_i, \text{BL}_i < \kappa_{\text{th}} \end{cases}. \tag{29}$$

2) When $\mu_i > Q_H$, that is, when the user is in a higher QoE, the buffer level threshold κ_{th} is calculated as shown in Eq. 30, where b determines the speed of user i to establish a buffer.

$$\kappa_{\text{th}} = \frac{\mu_i - Q_H}{b}. \tag{30}$$

Equation 30 shows that the higher the user's transmission rate or achievable MOS value, the greater the buffer level threshold. The faster the buffer is built, the lower the buffer level threshold. When the buffer level is below the threshold, it indicates that the buffer level capacity is low, making it difficult to support users to watch a higher bit rate version φ_i. When the buffer level is above the threshold, it is sufficient to remain below the lowest-resolution version at which the transfer rate can be achieved.

$$\begin{cases} Q_i(\varphi_i) \ge \mu_i, \text{BL}_i \ge \kappa_{\text{th}} \\ Q_i(\varphi_i) \le Q_H + b\,\text{BL}_i, \text{BL}_i < \kappa_{\text{th}} \end{cases}. \tag{31}$$

5.2 Joint bandwidth allocation and quality adaptation with buffer-level awareness

Flexible reconfiguration of the wireless bandwidth according to the user buffer level can further improve user satisfaction. For example, without reducing the bit rate version requested by the user, part of the bandwidth resources from users with high buffer levels can be allocated to support higher-resolution streams and buffer them for users with lower levels. This approach is different from determining the transmission and playback rates separately; as discussed in Section 5.2, the optimization problem (32) jointly configures the transmission and playback rates according to the buffer level to improve the user's QoE.

$$\begin{aligned} \delta_{\text{opt}} &= \arg\max_{\{\delta_1,\delta_2,\ldots,\delta_N\}} \frac{1}{N}\sum_{i=1}^{N}\{Q_i(\gamma_i, \text{BL}_i) - P_i\}, \\ \text{s.t. } &\sum_{i=1}^{N}\delta_i = 1. \end{aligned} \tag{32}$$

In this paper, a greedy algorithm is proposed to solve problem Eq. 32, and the optimal wireless resource allocation strategy is obtained through multiple iterations. First, using Eqs 23, 28–31, the wireless bandwidth configuration and resolution selection sets

TABLE 1 Simulation parameters.

Parameter	Value
Number of RRH–ONU	20
RRH coverage	500 × 500 m^2
Number of videos	10^4
Wavelength capacity	10 Gbps
Average wireless access delay	50 ms
Average fronthaul delay and average backhaul delay	20 and 80 ms
Transmission power	20 dBm
RRH cache capacity	100 GB
System bandwidth	20 MHz
Video resolution version	11
Path loss	$L = 128 + 37.61gR$ dB

are initialized, respectively, that is, $\{\gamma_1, \gamma_2, \ldots, \gamma_N\}^0, \{\varphi_1, \varphi_2, \ldots, \varphi_N\}^0$. The user's utility function is $U_i = Q_i(\gamma_i, \mathrm{BL}_i) - P_i$. Next, in each iteration process, users i and j are searched so that when increasing and decreasing their wireless share $\Delta\delta$, the difference between the utility increment and decrement of the two users is maximized, as provided in Eq. 33:

$$\arg\max_{i,j} \Delta U_{ij} = \Delta U_i\left(\mathrm{g}_i\left(\delta_i + \Delta\delta\right)\right) - \Delta U_j\left(\mathrm{g}_j\left(\delta_j - \Delta\delta\right)\right). \quad (33)$$

After confirming the user pair, update $(\gamma_i)^m$ and $(\varphi_i)^m$ and $(\gamma_j)^m$ and $(\varphi_j)^m$, respectively. When the utility increment ΔU_{ij} of the user pair is extremely small (i.e., $\Delta U_j \le \Delta U_{\min}$), stop the iteration and output $\{\delta_1, \delta_2, \ldots, \delta_N\}$. Edge users are provided with wireless resources by multiple RRHs; therefore, edge user k takes the minimum wireless bandwidth share of RRHs in its wireless domain serving cell, that is, $\delta_i = \min\{\delta_{ir}, r \in \mathrm{WSC}_k\}$. Each RRH will have remaining bandwidth resources, so the wireless bandwidth share of edge users is fixed, and resources are redistributed according to the above greedy algorithm.

6 Simulation results

This paper uses the NS2 simulation platform to verify the proposed QoE-aware video cooperative caching and transmission mechanism. The quality-aware wireless edge caching (QAWEC) algorithm is compared with the RARCP algorithm proposed by [54]. The CEC algorithm selects an appropriate RRH set for edge users based on the static cache information and provides them with joint transmission services to ensure that the user minimizes the backhaul traffic under the condition of the wireless link rate. The QAWEC algorithm establishes a joint optimization problem of bandwidth configuration and edge cache according to user behavior to ensure the cache hit rate and user experience quality. The RARCP algorithm formulates a user preference (UPP, user preference profile)-aware active pre-caching strategy and then schedules backhaul and wireless resources based on the cached information, thereby maximizing user QoE. The simulation parameter settings are shown in Table 1.

6.1 Performance evaluation under different number of users

Edge user throughput is one of the key factors in measuring user QoE, which is defined as the amount of data successfully transmitted by the wireless domain serving cell for edge user joint transmission per unit time. Figure 3 shows the relationship between the number of different users and the average throughput of edge users. It can be seen from Figure 3 that, with the increase in the number of users, the average throughput of edge users shows a downward trend. The

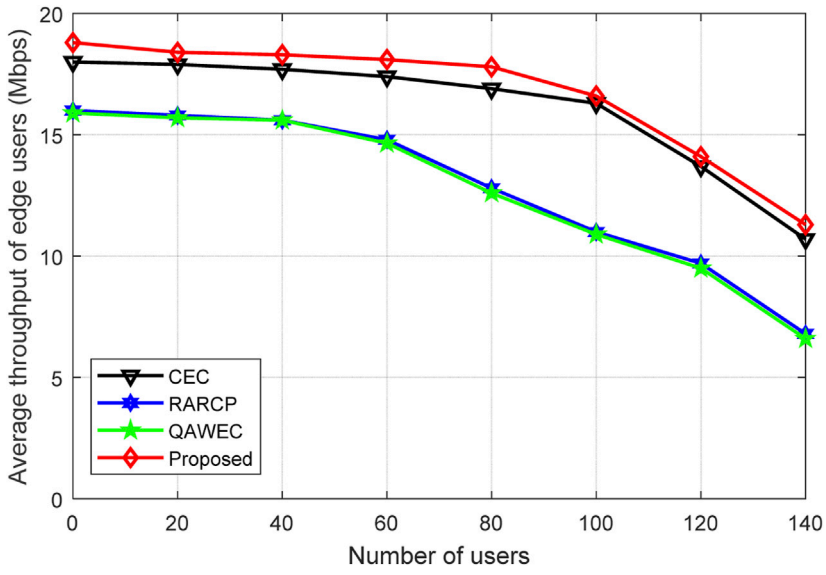

FIGURE 3
Relationship between the number of different users and the average throughput of edge users.

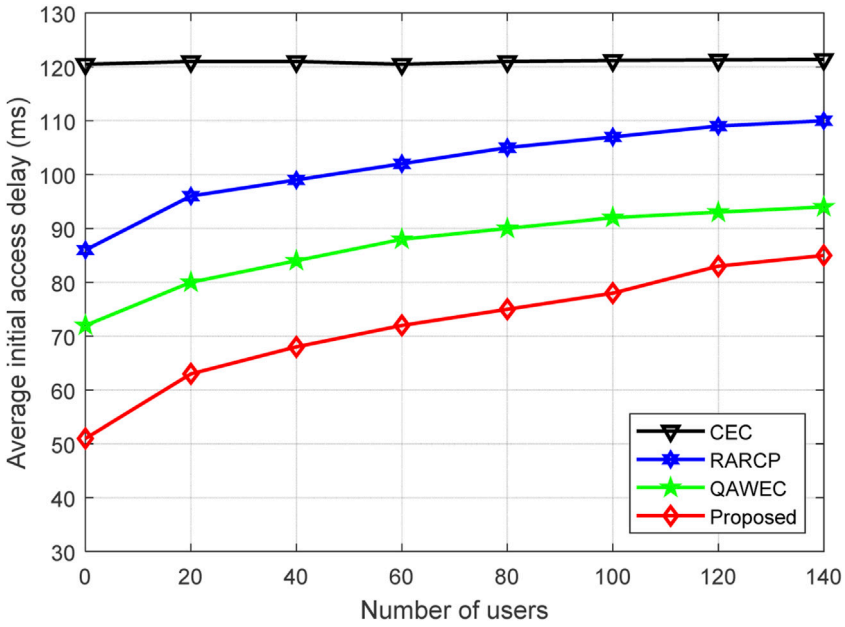

FIGURE 4
Relationship between the number of different users and the average initial access delay of users.

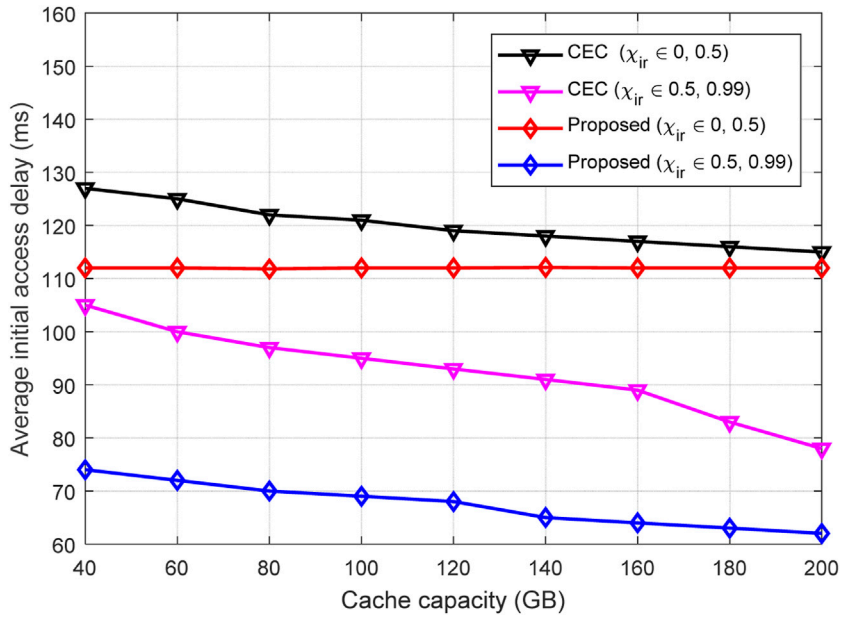

FIGURE 5
Relationship between the cache capacity and average initial access delay under different user behaviors.

performance of the proposed QoE-aware video cooperative caching and transmission mechanism (QACCTM) algorithm is about 30% higher than that of other algorithms. This is mainly because the cooperative caching and transmission gain of the wireless domain are considered in the construction of VPON so that edge users can use the cached content in the serving cell of the wireless domain to provide joint data for them and do not need to fetch cache from content server. This not only enhances the data link rate of edge users but also avoids the packet loss rate caused by video transmission from content servers, thereby increasing the throughput of edge users. However, the joint transmission cluster constructed by the CEC algorithm makes some RRHs still need to obtain the cache through the content server, so its performance is slightly lower than that of the proposed QACCTM algorithm. The QAWEC and RARCP algorithms have poor throughput because they do not consider the performance characteristics of edge users.

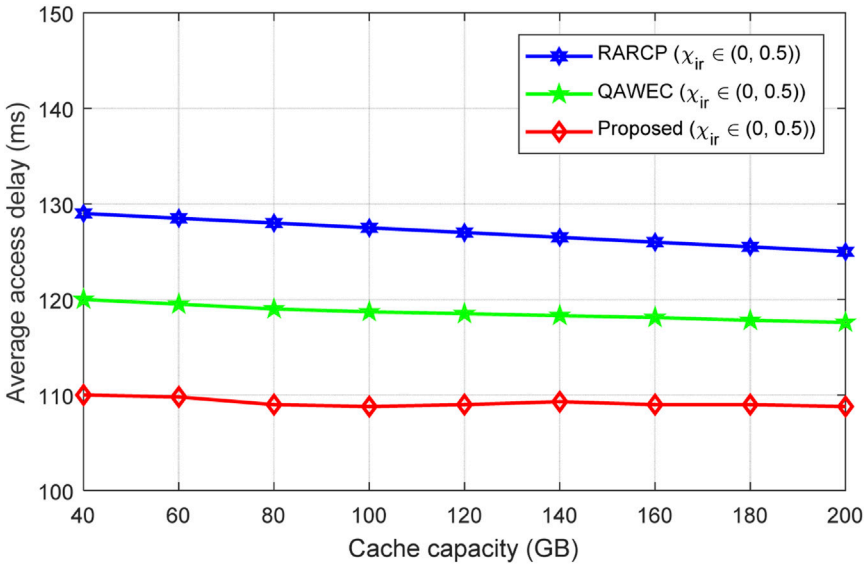

FIGURE 6
Relationship between the cache capacity and average initial access delay under different algorithms.

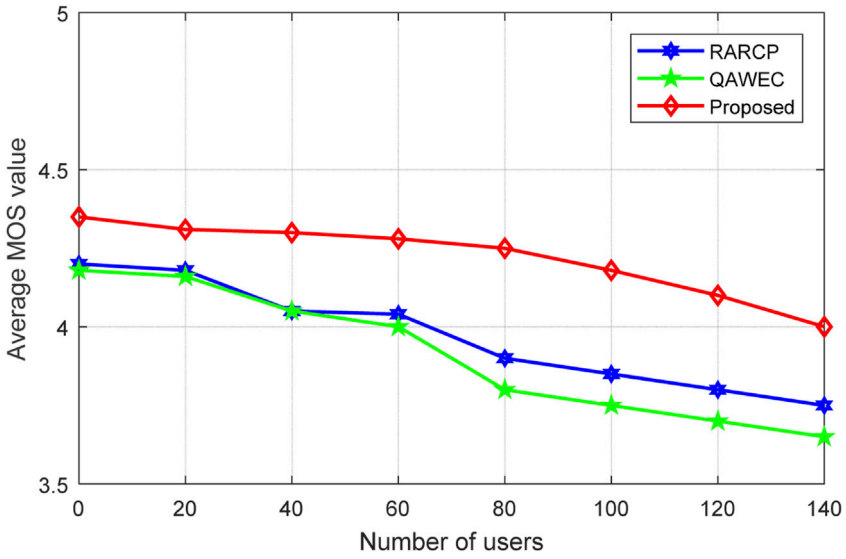

FIGURE 7
Relationship between the number of different users and the average MOS value of users.

Figure 4 reflects the relationship between the number of different users and the average initial access delay of users. It can be seen from Figure 4 that with the increase in the number of users, the average initial access delay of users tends to increase. However, due to the invariance of the cache content, the CEC algorithm makes the change in the number of users have little effect on the average access delay of users, and the resource utilization rate is low. The RARCP algorithm dynamically updates the cache through the user's content preference, ignoring the importance of the user's mobile mode for the cache hit rate. The QAWEC algorithm does not pay attention to the difference in the cache hit rate between light and heavy load cells, only optimizes the cached content of a single cell, and does not effectively utilize the cache resources of light load cells. However, the proposed QACCTM algorithm comprehensively considers the user's behavioral characteristics (request preference and mobile mode) and the cooperative caching and transmission of the wireless and optical domains and reduces the access of heavy-load cell and edge users by constructing optical domain or wireless domain service cells. Therefore, compared with the RARCP and QAWEC algorithms, the proposed QACCTM algorithm reduces the access delay by approximately 27.1% and 15.9%, respectively.

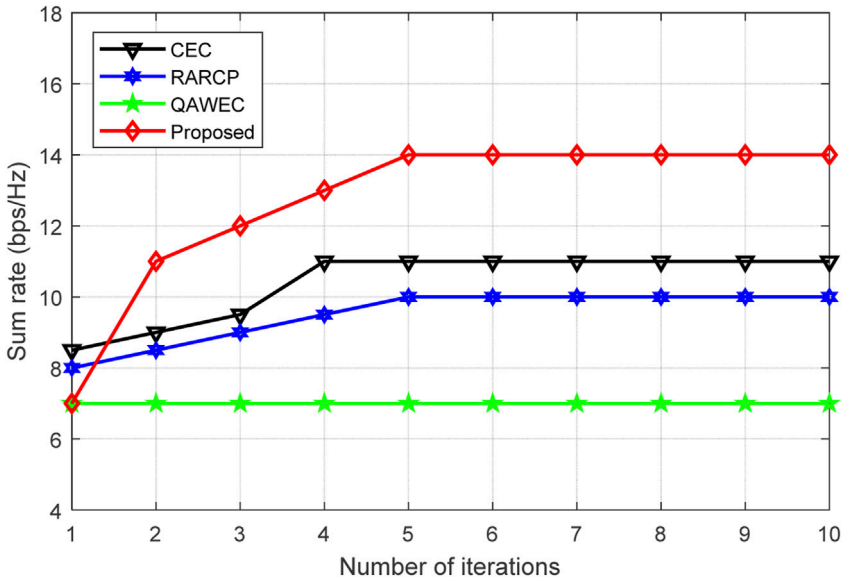

FIGURE 8
Sum rate comparison of the algorithms with an increasing number of iterations.

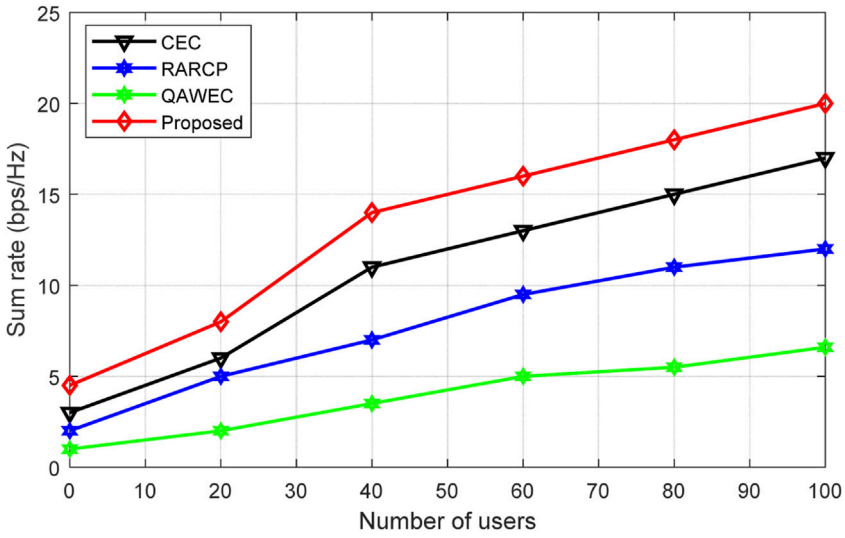

FIGURE 9
Sum rate comparison of algorithms with an increasing number of users.

6.2 Performance evaluation under different cache capacity values

Cache capacity is an important factor affecting user access latency. When the probability $\chi_{ir} \in (0.5,0.99]$ of all users requesting cached content in the network is high, user behavior is easier to obtain. When $\chi_{ir} \in (0,0.5]$, user behavior is difficult to capture. Figure 5 shows the relationship between the cache capacity and average initial access delay under different user behaviors. It can be seen from Figure 5 that in the scenario of $\chi_{ir} \in (0.5,0.99]$, the average access delay of the user is shorter because the user's behavior is easier to obtain, and the hit rate of dynamic pre-caching is higher. Since the CEC algorithm is in the static cache mode, as the cache capacity increases, its access delay tends to decrease slightly. However, the access delay of the proposed QACCTM algorithm is not sensitive to the cache capacity because the dynamic pre-caching according to the user's behavior characteristics greatly improves the utilization of the cache space.

Figure 6 reflects the relationship between the cache capacity and the average initial access delay under different algorithms. It can be seen from Figure 6 that the access delay of the RARCP algorithm is relatively high. The reason is that in the scenario of $\chi_{ir} \in (0,0.5]$, it is difficult to

capture user behavior and the cache hit rate is low. Compared with the QAWEC and QACCTM algorithms, the RARCP algorithm is more sensitive to the cache capacity because the increase in the cache capacity makes the user's cache hit rate higher in this scenario. However, the QAWEC and QACCTM algorithms are not sensitive to the cache capacity because the two algorithms take into account the user's behavior characteristics for pre-caching, which improves the cache hit rate. Therefore, even if the cache capacity increases, the increased cache space utilization rate is not high. In addition, the proposed algorithm potentially increases the buffer capacity of each RRH through the cooperative buffering of the optical and wireless domains and can achieve better performance even when the buffer capacity is low.

6.3 Performance evaluation of user QoE

Video quality is a key factor in evaluating user QoE. Figure 7 shows the relationship between the number of different users and the average MOS value of users. Since the CEC algorithm does not consider the impact of wireless resource allocation on the user's viewing experience, this paper only makes a comparative analysis of the other three algorithms. Figure 7 shows that with an increase in the number of users, the average MOS value of users shows a downward trend. The performance of the QAWEC and RARAP algorithms is poor because as the user scale increases in order to avoid video interruption, both algorithms will greatly reduce the video resolution version that users watch. However, the performance of the proposed QACCTM algorithm is 5.6% higher than that of other algorithms. This algorithm flexibly configures wireless bandwidth resources through the buffer level and enables users with high buffer levels to request higher bit rates while ensuring continuous video playback. In the high-resolution version, users with a low buffer level can call the bandwidth resources of users with a high buffer level to fill the buffer and maintain the resolution version currently played by the user.

The smoothness of video playback is also an important parameter to measure user QoE. If the video resolution version is too different in adjacent scheduling periods, it will cause non-smooth playback of the video. During video playback, the more the instances of non-smooth switching, the worse the smoothness of the video, which will reduce the user QoE. In this paper, quality switching over one resolution level is regarded as a measure of non-smooth switching.

Figure 8 compares the sum rate of the algorithms under different numbers of iterations. Figure 8 shows that the sum rate of the proposed algorithm is higher with an increasing number of iterations than that of existing algorithms. This further validates the applicability of the proposed algorithm in large usage scenarios and its ability to provide better QoS.

Figure 9 compares the sum rate of the algorithms when the number of users increases. Figure 9 shows that the sum rate of the proposed algorithm is higher than that of existing algorithms, which validates its effectiveness.

7 Conclusion

In order to make full use of network resources and improve the buffering and transmission performance of multimedia services, this paper proposes a QoE-aware video cooperative buffering and transmission mechanism based on the joint architecture of TWDM-PON and C-RAN. First, the degree of interference between cells, the similarity of request content, and the transmission bandwidth of ONUs are used to represent the cooperative buffering and transmission gain of the optical and wireless domains, and a cooperatively aware virtual passive optical network is formed through RRH and ONU clustering. Furthermore, the user's video experience, bandwidth configuration, and caching strategy are jointly optimized. On the optical side, a joint fronthaul bandwidth and buffer allocation mechanism is proposed to reduce user access delay through dynamic pre-caching. On the wireless side, the flexible configuration of wireless bandwidth is realized according to the buffer level, which enhances the quality of video watched by users. The simulation results show that the proposed mechanism effectively improves the cache hit rate while enhancing the quality of the user experience. In the future study, dynamic resource allocation in the heterogeneous network will be considered with different parametric analyses.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material; further inquiries can be directed to the corresponding author/s.

Author contributions

JT: conceptualization, data curation, resources, software, validation, and writing–original draft. HC: data curation, investigation, project administration, resources, supervision, and writing–original draft. JC: data curation, methodology, project administration, resources, software, and writing–original draft. IK: conceptualization, data curation, methodology, project administration, validation, and writing–original draft. AA: formal analysis, investigation, methodology, validation, visualization, and writing–original draft. P-CW: formal analysis, methodology, resources, software, visualization, and writing–original draft. IH: funding acquisition, investigation, methodology, project administration, resources, supervision, and writing–original draft.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. National Science and Technology Council 112-2811-E-005-011-MY2, 112-2634-F-005-001-MBK and 112-2634-F-005-002.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Liu H, Wang H, Ji Y. Simultaneous all-optical channel aggregation and de-aggregation for 8QAM signal in elastic optical networking. *IEEE Photon J* (2019) 11(1):1–8. doi:10.1109/JPHOT.2018.2884763

2. Klinkowski M, Jaworski M. Cost-aware optimization of optical add-drop multiplexers placement in packet-optical xHaul access networks. *Appl Sci* (2023) 13(8):4862–15. doi:10.3390/app13084862

3. Alrubayyi H, Alshareef M, Nadeem Z, Abdelmoniem A, Jaber M. Security threats and promising solutions arising from the intersection of AI and IoT: a study of IoMT and IoET applications. *Future Internet* (2024) 16(3):85–18. doi:10.3390/fi16030085

4. Alenizi F, Rana O. Dynamically controlling offloading thresholds in fog systems. *Sensors* (2021) 21(7):2512–5. doi:10.3390/s21072512

5. Suo L, Qi L, Wang L. Link load correlation-based blocking performance analysis for tree-type data center networks. *Appl Sci* (2022) 12(12):6235–14. doi:10.3390/app12126235

6. Bonani L, Forghani M. An improved least cost routing approach for WDM optical network without wavelength converters. *Opt Fiber Tech* (2016) 16(32):30–5. doi:10.1016/j.yofte.2016.09.003

7. Kaur H, Rattan M. Improved offline multi-objective routing and wavelength assignment in optical networks. *Front Optoelectronics* (2019) 12:433–44. doi:10.1007/s12200-019-0850-4

8. Abdo A, Amours C. Adaptive pre/post-compensation of cascade filters in coherent optical transponders. *Future Internet* (2020) 12(2):1–19. doi:10.3390/fi12020021

9. Zheng Z, Li M, Tseng T, Schlichtmann U. LightR: a fault-tolerant wavelength-routed optical network-on-chip topology. *Appl Sci* (2023) 13(15):1–14. doi:10.3390/app13158871

10. Wang Y, Li C, Hu Q, Flor J, Jalalitabar M. Routing and spectrum allocation in spectrum-sliced elastic optical networks: a primal-dual framework. *Electronics* (2021) 10(22):1–22. doi:10.3390/electronics10222809

11. Virgillito E, Ferrari A, Damico A, Curri V. Statistical assessment of open optical networks. *Photonics* (2019) 6(2):64–15. doi:10.3390/photonics6020064

12. Yu X, Ning X, Zhu Q, Lv J, Zhao Y, Zhang H, et al. Multi-dimensional routing, wavelength, and timeslot allocation (RWTA) in quantum key distribution optical networks (QKD-ON). *Appl Sci* (2021) 11(1):348–18. doi:10.3390/app11010348

13. Wu J, Subramanian S, Hasegawa H. Efficient dynamic routing and spectrum assignment for multifiber elastic optical networks. *J Opt Commun Networking* (2019) 11:190–201. doi:10.1364/JOCN.11.000190

14. Zhai Z, Dou L, He Y, Lau A, Xie C. Open-source data for QoT estimation in optical networks from Alibaba. *J Opt Commun Networking* (2024) 16(1):1–3. doi:10.1364/JOCN.504549

15. Ricciardi S, Sembroiz D, Palimieri F, Santos-Boada G, Perelló J, Careglio D. A hybrid load-balancing and energy-aware RWA algorithm for telecommunication networks. *Comput Commun* (2015) 77(3):85–99. doi:10.1016/j.comcom.2015.06.010

16. Pavarangkoon P, Oki E. A routing and wavelength assignment scheme considering full optical carrier replication in multi-carrier-distributed optical mesh networks with wavelength reuse. *Opt Switching Networking* (2018) 18(28):23–35. doi:10.1016/j.osn.2017.12.001

17. Hsu C, Cho H, Fang S. Solving routing and wavelength assignment problem with maximum edge-disjoint paths. *J Ind Manag Optimization* (2017) 13(2):1065–84. doi:10.3934/jimo.2016062

18. Muro F, Garrich M, Castreno I, Zahir S, Marino P. Emulating software-defined disaggregated optical networks in a containerized framework. *Appl Sci* (2021) 11(5):1–17. doi:10.3390/app11052081

19. Jha R, Llah B. Software-defined optical networks (SDON): proposed architecture and comparative analysis. *J Eur Opt Society-Rapid Publications* (2019) 15(16):16–3. doi:10.1186/s41476-019-0105-4

20. Zhang S, Xue X, Tangdiongga E, Calabretta N. Low-latency optical wireless data-center networks using nanoseconds semiconductor-based wavelength selectors and arrayed waveguide grating router. *Photonics* (2022) 9(3):203–17. doi:10.3390/photonics9030203

21. Wang P, Yang L, Nie X, Ren Z, Li J, Kuang L. Data-driven software defined network attack detection: state-of-the-art and perspectives. *Inf Sci* (2020) 513:65–83. doi:10.1016/j.ins.2019.08.047

22. Yang P, Chen L, Zhang H, Yang J, Wang R, Li Z. Joint optical and wireless resource allocation for cooperative transmission in C-RAN. *Sensors* (2021) 21(1):217–8. doi:10.3390/s21010217

23. Wang B, Peng L, Ho P. Energy-efficient radio-over-fiber system for next-generation cloud radio access networks. *EURASIP J Wireless Commun Networking* (2019) 118:118–21. doi:10.1186/s13638-019-1457-6

24. Dryjanski M, Kulacz L, Kliks A. Toward modular and flexible open RAN implementations in 6G networks: traffic steering use case and O-RAN xApp. *Sensors* (2021) 21(24):1–24. doi:10.3390/s21248173

25. He S, Tian H, Lyn X, et al. Distributed cache placement and user association in multicast-aided heterogeneous network. *IEEE Access* (2017) 5(3):25365–76. doi:10.1109/ACCESS.2017.2769664

26. Hou T, Feng G, Qin S, Jiang W. Proactive content caching by exploiting transfer learning for mobile edge computing. *Int J Commun Syst* (2018) 31(11):3706–16. doi:10.1002/dac.3706

27. Argyrious A, Poularakis K, Iosifidis G, Tassiulas L. Video delivery in dense 5G cellular networks. *IEEE Netw* (2017) 31(4):28–34. doi:10.1109/MNET.2017.1600298

28. Xiao T, Cui T, Islam S, Chen Q. Joint content placement and storage allocation based on federated learning in F-RANs. *Sensors* (2021) 21(1):215–7. doi:10.3390/s21010215

29. Li Y. Federated deep reinforcement learning-based caching and bitrate adaptation for VR panoramic video in clustered MEC networks. *Electronics* (2022) 11(23):3968–14. doi:10.3390/electronics11233968

30. Rezaee A, Sheikhabad O, Beygi L. Quality of transmission-aware control plane performance analysis for elastic optical networks. *Computer Networks* (2021) 187(3):7755–68. doi:10.1016/j.comnet.2020.107755

31. Padmaloshani P, Nirmala S. Semi-distributed dynamic inter-cell interference coordination scheme for interference avoidance in heterogeneous networks. *ETRI J* (2020) 42(2):175–85. doi:10.4218/etrij.2018-0362

32. Hou R, Cai J, Lui K. Distributed cache-aware CoMP transmission scheme in dense small cell networks with limited backhaul. *Comput Commun* (2019) 138(3):11–9. doi:10.1016/j.comcom.2018.12.004

33. Lyko T, Broadbent M, Race N, Nilsson M, Farrow P, Appleby S. Improving quality of experience in adaptive low latency live streaming. *Multimedia Tools Appl* (2023) 8(3):15957–83. doi:10.1007/s11042-023-15895-9

34. Tran T, Hajisami A, Pompili P. Cooperative hierarchical caching in 5G cloud radio access networks. *IEEE Netw* (2017) 31(4):35–41. doi:10.1109/MNET.2017.1600307

35. Khatibi S, Caeiro L, Ferreira L, Correia L, et al. Modeling and implementation of virtual radio resources management for 5G cloud RAN. *EURASIP J Wireless Commun Networking* (2017) 128:1–14. doi:10.1186/s13638-017-0908-1

36. Yin Y, Guo Y, Su Q, Wang Z. Task allocation of multiple unmanned aerial vehicles based on deep transfer reinforcement learning. *Drones* (2022) 6(8):215. doi:10.3390/drones6080215

37. Xu X, Liu W, Yu L. Trajectory prediction for heterogeneous traffic-agents using knowledge correction data-driven model. *Inf Sci* (2022) 608:375–91. doi:10.1016/j.ins.2022.06.073

38. Yan A, Chen Y, Gao Z, Ni T, Huang Z, Cui J, et al. FeMPIM: a FeFET-based multifunctional processing-in-memory cell. *IEEE Trans Circuits Syst Express Briefs* (2024) 71(4):2299–303. doi:10.1109/TCSII.2023.3331267

39. Liu J, Wang H, Liu M, Zhao R, Zhao Y, Sun T, et al. POMDP-based real-time path planning for manipulation of multiple microparticles via optoelectronic tweezers. *Cyborg Bionic Syst* (2022) 2022:9890607. doi:10.34133/2022/9890607

40. Xu G, Zhang Q, Song Z, Ai B. Relay-assisted deep space optical communication system over coronal fading channels. *IEEE Trans Aerospace Electron Syst* (2023) 59(6):8297–312. doi:10.1109/TAES.2023.3301463

41. Cheng D, Chen L, Lv C, Guo L, Kou Q. Light-guided and cross-fusion U-net for anti-illumination image super-resolution. *IEEE Trans Circuits Syst Video Tech* (2022) 32(12):8436–49. doi:10.1109/TCSVT.2022.3194169

42. Fu C, Yuan H, Xu H, Zhang H, Shen L. TMSO-Net: texture adaptive multi-scale observation for light field image depth estimation. *J Vis Commun Image Representation* (2023) 90:103731. doi:10.1016/j.jvcir.2022.103731

43. Liang X, Zhao Y, Liu D, Deng Y, Arai T, Kojima M, et al. Magnetic microrobots fabricated by photopolymerization and assembly. *Cyborg Bionic Syst* (2023) 4:0060. doi:10.34133/cbsystems.0060

44. Yang M, Cai C, Wang D, Wu Q, Liu Z, Wang Y. Symmetric differential demodulation-based heterodyne laser interferometry used for wide frequency-band

vibration calibration. *IEEE Trans Ind Electro* (2024) 71(7):8132–40. doi:10.1109/TIE.2023.3299015

45. Wu Z, Ismail M. Generalized RIS tile exclusion strategy for indoor mmWave channels under concept drift. *IEEE Trans Wireless Commun* (2024) 1. doi:10.1109/TWC.2024.3402267

46. Wang Y, Xiao R, Xiao N, Wang Z, Chen L, Wen Y, et al. Wireless multiferroic memristor with coupled giant impedance and artificial synapse application. *Adv Electron Mater* (2022) 8(10):2200370. doi:10.1002/aelm.202200370

47. Liu D, Cao Z, Jiang H, Zhou S, Xiao Z, Zeng F. Concurrent low-power listening: a new design paradigm for duty-cycling communication. *ACM Trans Sen Netw* (2022) 19(1):1–24. doi:10.1145/3517013

48. Dai M, Luo L, Ren J, Yu H, Sun G. PSACCF: prioritized online slice admission control considering fairness in 5G/B5G networks. *IEEE Trans Netw Sci Eng* (2022) 9(6):4101–14. doi:10.1109/TNSE.2022.3195862

49. Sun G, Xu Z, Yu H, Chang V. Dynamic network function provisioning to enable network in box for industrial applications. *IEEE Trans Ind Inform* (2021) 17(10):7155–64. doi:10.1109/TII.2020.3042872

50. Sun G, Sheng L, Luo L, Yu H. Game theoretic approach for multipriority data transmission in 5G vehicular networks. *IEEE Trans Intell Transportation Syst* (2022) 23(12):24672–85. doi:10.1109/TITS.2022.3198046

51. Sun G, Song L, Yu H, Chang V, Du X, Guizani M. V2V routing in a VANET based on the autoregressive integrated moving average model. *IEEE Trans Vehicular Tech* (2019) 68(1):908–22. doi:10.1109/TVT.2018.2884525

52. He T, Zheng Y, Liang X, Li J, Lin L, Zhao W, et al. A highly energy-efficient body-coupled transceiver employing a power-on-demand amplifier. *Cyborg Bionic Syst* (2023) 4:0030. doi:10.34133/cbsystems.0030

53. Chen B, Hu J, Ghosh BK. Finite-time tracking control of heterogeneous multi-AUV systems with partial measurements and intermittent communication. *SCIENCE CHINA Inf Sci* (2024) 67(5):152202. doi:10.1007/s11432-023-3903-6

54. Liu Z, Qiao B, Fang K. Joint optimization strategy for QoE-aware encrypted video caching and content distributing in multi-edge collaborative computing environment. *J Cloud Comput* (2020) 56:56–19. doi:10.1186/s13677-020-00204-8

55. Hu J, Wu Y, Li T, Ghosh BK. Consensus control of general linear multiagent systems with antagonistic interactions and communication noises. *IEEE Trans Automatic Control* (2019) 64(5):2122–7. doi:10.1109/TAC.2018.2872197

56. Chen B, Hu J, Zhao Y, Ghosh BK. Finite-time velocity-free rendezvous control of multiple AUV systems with intermittent communication. *IEEE Trans Syst Man, Cybernetics: Syst* (2022) 52(10):6618–29. doi:10.1109/TSMC.2022.3148295

57. Tang Q, Qu S, Zhang C, Tu Z, Cao Y. Effects of impulse on prescribed-time synchronization of switching complex networks. *Neural Networks* (2024) 174:106248. doi:10.1016/j.neunet.2024.106248

58. Hao J, Chen P, Chen J, Li X. Multi-task federated learning-based system anomaly detection and multi-classification for microservices architecture. *Future Generation Comput Syst* (2024) 159:77–90. doi:10.1016/j.future.2024.05.006

59. Xuemin Z, Haitao D, Zenggang X, Ying R, Yanchao L, Yuan L, et al. Self-organizing key security management algorithm in socially aware networking. *J Signal Process Syst* (2024) 96:369–83. doi:10.1007/s11265-024-01918-7

60. Zhou G, Xu C, Zhang H, Zhou X, Zhao D, Wu G, et al. PMT gain self-adjustment system for high-accuracy echo signal detection. *Int J Remote Sensing* (2022) 43(19-24):7213–35. doi:10.1080/01431161.2022.2155089

61. Zhou G, Zhang H, Xu C, Zhou X, Liu Z, Zhao D, et al. A real-time data acquisition system for single-band bathymetric LiDAR. *IEEE Trans Geosci Remote Sensing* (2023) 61:1–21. doi:10.1109/TGRS.2023.3282624

62. Zhou G, Zhao D, Zhou X, Xu C, Liu Z, Wu G, et al. An RF amplifier circuit for enhancement of echo signal detection in bathymetric LiDAR. *IEEE Sensors J* (2022) 22(21):20612–25. doi:10.1109/JSEN.2022.3206763

63. Yang Y, Wei X, Yao W, Lan J. Broadband electrical impedance matching of sandwiched piezoelectric ultrasonic transducers for structural health monitoring of the rail in-service. *Sensors Actuators A: Phys* (2023) 364:114819. doi:10.1016/j.sna.2023.114819

64. Zhu C, Al-Dossari M, Rezapour S, Alsallami SAM, Gunay B. Bifurcations, chaotic behavior, and optical solutions for the complex Ginzburg–Landau equation. *Results Phys* (2024) 59:107601. doi:10.1016/j.rinp.2024.107601

65. Zhu C, Al-Dossari M, Rezapour S, Shateyi S, Gunay B. Analytical optical solutions to the nonlinear Zakharov system via logarithmic transformation. *Results Phys* (2024) 56:107298. doi:10.1016/j.rinp.2023.107298

66. Hui Z, Wu A, Han D, Li T, Li L, Gong J, et al. Switchable single-to multiwavelength conventional soliton and bound-state soliton generated from a NbTe2 saturable absorber-based passive mode-locked erbium-doped fiber laser. *ACS Appl Mater Inter* (2024) 16(17):22344–60. doi:10.1021/acsami.3c19323

67. Wen C, Huang Y, Peng J, Wu J, Zheng G, Zhang Y. Slow-time FDA-MIMO technique with application to STAP radar. *IEEE Trans Aerospace Electron Syst* (2022) 58(1):74–95. doi:10.1109/TAES.2021.3098100

68. Wang R, Gu Q, Lu S, Tian J, Yin Z, Yin L, et al. FI-NPI: exploring optimal control in parallel platform systems. *Electronics* (2024) 13(7):1168. doi:10.3390/electronics13071168

69. Zhang R, Cheng L, Wang S, Lou Y, Gao Y, Wu W, et al. Integrated sensing and communication with massive mimo: a unified tensor approach for channel and target parameter estimation. *IEEE Trans Wireless Commun* (2024) 1. doi:10.1109/TWC.2024.3351856

70. Fu Y, Doan K, Quek T. On recommendation-aware content caching for 6G: an artificial intelligence and optimization empowered paradigm. *Digital Commun Networks* (2020) 6(3):304–11. doi:10.1016/j.dcan.2020.06.005

71. Wang Q, Dai H, Wu D, Xiao H. Data analysis on video streaming QoE over mobile networks. *EURASIP J Wireless Commun Networking* (2018) 173:173–21. doi:10.1186/s13638-018-1180-8

72. Katsenou A, Afonso M, Bull D. Study of compression statistics and prediction of rate-distortion curved for video texture. *Signal Processing: Image Commun* (2022) 101(3):3312–24. doi:10.1016/j.image.2021.116551

73. Liang C, He Y, Yu R, Zhao N. Enhancing QoE-aware wireless edge caching with software-defined wireless networks. *IEEE Trans Wireless Commun* (2017) 16(10):6912–25. doi:10.1109/TWC.2017.2734081