

基于 TWDM-PON 与 C-RAN 的 QoE 感知视频 协作缓存与传输机制

吴大鹏^{1,2}, 李雪^{1,2}, 李红霞³

(1. 重庆邮电大学通信与信息工程学院, 重庆 400065; 2. 重庆市光通信与网络重点实验室, 重庆 400065;
3. 中国联合网络通信有限公司重庆市分公司, 重庆 401123)

摘 要: 随着数据流量的爆发式增长, 用户体验质量急剧恶化, 为了解决密集组网与负载不均衡所引发的小区边缘用户与重负载小区用户性能较差的问题, 提出了一种体验质量感知的云无线接入网视频协作缓存与传输机制。该机制通过建立协作增益感知的虚拟无源光网络, 采用光域和无线域协同的方式为视频内容提供协作缓存与传输; 进而, 联合优化用户视频体验、带宽配置与缓存策略, 分别在光域和无线域采用动态预缓存及缓冲水平感知的带宽配置方式, 达到提高用户体验质量的目的。结果表明, 所提机制能够有效地提高缓存命中率, 增强用户体验质量。

关键词: 云无线接入网; 虚拟无源光网络; 协作缓存与传输; 体验质量

中图分类号: TP302

文献标识码: A

doi: 10.11959/j.issn.1000-436x.2019066

QoE-aware video cooperative caching and transmission mechanism based on TWDM-PON and C-RAN

WU Dapeng^{1,2}, LI Xue^{1,2}, LI Hongxia³

1. School of Telecommunication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
2. Key Laboratory of Optical Communication and Network, Chongqing 400065, China
3. Chongqing Branch of China Unicom, Chongqing 401123, China

Abstract: The quality of user experience suffers from performance deterioration dramatically due to the explosively growing data traffic. To improve the poor performance of cell-edge users and heavy-load cell users, which caused by dense network and load imbalance respectively, an QoE-aware video cooperative caching and transmission mechanism in cloud radio access network was proposed. Cooperative gain-aware virtual passive optical network was established to provide cooperative caching and transmission for video streaming by adopting collaborative approach in optical domain and wireless domain. Furthermore, user experience for video streaming, bandwidth provisioning and caching strategy were jointly optimized to improve QoE, which utilized the methods of dynamic caching in optical domain and buffer level-aware bandwidth configuration in wireless domain. The results show that the proposed mechanism enhances the quality of user experience and effectively improves the cache hit rate.

Key words: cloud radio access network, virtual passive optical network, cooperative caching and transmission, quality of experience

收稿日期: 2018-04-28; 修回日期: 2018-09-20

基金项目: 国家自然科学基金资助项目 (No.61771082, No.61871062); 重庆市高校创新团队建设计划基金资助项目 (No.CXTDX201601020)

Foundation Items: The National Natural Science Foundation of China (No.61771082, No.61871062), Chongqing Funded Project of Chongqing University Innovation Team Construction (No.CXTDX201601020)

1 引言

随着多媒体智能终端设备的广泛应用，视频业务呈现爆发式增长趋势，传统的分布式无线接入网（D-RAN, distributed radio access network）无法满足不断增长的用户需求与网络需求。为了解决上述问题，研究人员提出了基于集中化处理、协作式无线电和实时云计算的云无线接入网（C-RAN, cloud radio access network）架构^[1]。与 D-RAN 不同，C-RAN 打破了远端无线射频单元（RRH, remote radio head）和基带处理单元（BBU, base band unit）之间的固定连接关系，使 RRH 分布在各小区站点，BBU 集中在中心主机房^[2]，BBU 与 RRH 之间构成前传（fronthaul）。显然，前传容量对 C-RAN 所承载的业务至关重要。由于时分波分复用无源光网络（TWDM-PON, time and wavelength division multiplexed passive optical network）具有高效、低时延及高传输容量的特点，能够满足视频业务的高带宽与时延敏感特性，从而成为 C-RAN 光纤前传的技术选择^[3]。

研究表明，用户点击视频的次数与视频流行度密切相关，且两者服从 Zipf 分布^[4]。大量用户不断地从内容服务器获取流行度高的视频，使网络所承载的数据流量变得异常庞大，导致用户体验质量（QoE, quality of experience）急剧下降，为此，研究高效的无线边缘缓存策略来减少数据冗余传输以提升移动网络性能势在必行^[5]。对于视频业务而言，访问时延与无线链路速率是影响用户体验质量的主要因素^[6-7]，因此，如何对视频内容进行缓存与传输对提高用户 QoE 至关重要。目前，普遍采用 RRH 边缘缓存的方式为用户提供视频传输，同时降低用户的访问时延。文献[8]在 C-RAN 架构下提出联合边缘缓存与云缓存策略，降低了平均时延，提高了用户 QoE。在非流量高峰时期为用户缓存视频内容，减轻预缓存占用带宽的压力，但由于用户的移动模式与请求内容都具有较强的动态特征，静态的预缓存策略无法满足用户的时变需求。文献[9]提出在软件定义的正交频分复用无源光网络中联合配置带宽与缓存资源，将光线路终端（OLT, optical line terminal）的部分带宽分配给时延敏感的流行视频，剩余带宽用于光网络单元（ONU, optical network unit）预缓存下一时隙的流行视频，从而降低服务响应时延，但该方案并未考虑超密集部署场景下的

用户特征^[10]，即小区边缘用户受到严重的同层干扰，并且小区负载不均衡会导致重负载小区用户性能较差。针对小区密集部署场景，文献[11]提出利用基站已有缓存为边缘用户进行联合传输分簇以最小化回传数据量，然而，在给定缓存资源下考虑联合传输无法保障边缘用户的 QoE，且缓存利用率低。文献[12]通过考虑视频段间的质量差异和用户缓冲区的信息来确定视频流的分辨率版本，从而在不中断播放的情况下保证单个用户的最佳观看质量，但这仅优化单个视频流的 QoE，当多个用户在共享链路上竞争资源时，会出现不公平的无线资源分配。

针对上述不足，在基于 TWDM-PON 与 C-RAN 的联合架构下，本文提出了一种 QoE 感知的视频协作缓存与传输机制。利用小区间干扰程度、请求内容相似度、ONU 传输带宽表征光与无线域协作缓存与传输增益，通过 RRH 与 ONU 分簇构建虚拟无源光网络（VPON, virtualized passive optical network）。进而，联合优化视频流 QoE、带宽配置与缓存分配：在光域，通过分析并计算视频的缓存价值与离散度动态地预缓存视频内容，降低访问时延；在无线域，基于各用户的缓冲水平灵活地配置无线带宽，提升用户观看视频的质量。

2 光域与无线域的协作架构

本文提出的基于 TWDM-PON 与 C-RAN 的联合架构如图 1 所示。光网络采用点到多点的树形结构，由 OLT 与多个 ONU 组成，无线网络由融合 ONU 与射频功能的网关节点（ONU-RRH, optical network unit-remote radio head）组成。

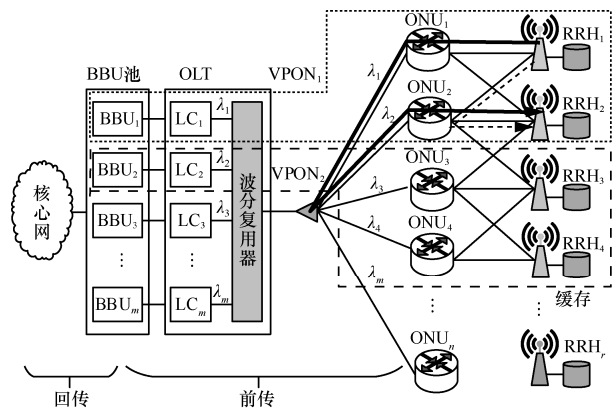


图 1 基于 TWDM-PON 与 C-RAN 的联合架构

与其他现有架构不同，本文提出的架构中每个 ONU 可以连接多个 RRH。BBU 池与 ONU-RRH 之

间的连接构成前传，与核心网的连接构成回传。基于 TWDM-PON 强大传输容量的特性，多个 ONU 可以利用其波长调谐器调谐到同一波长上进行波长共享。在光线路终端，由线卡 (LC, line card) 控制波长的接入。LC 与在其控制波长上通信的多个 ONU-RRH 的组合被称为虚拟 PON (VPON)，如图 1 所示，短虚线与长虚线框分别组成 VPON₁ 与 VPON₂。可见，利用 TWDM-POM 的波分特性可以建立多个 VPON，通过波分复用器进行光耦合与分离。

为提高小区边缘与重负载小区用户的 QoE，本文提出光域与无线域的协作缓存与传输，主要通过构建光域服务小区 (OSC, optical service cell) 与无线域服务小区 (WSC, wireless service cell) 来实现。无线域服务小区是以边缘用户为中心的 RRH 集合，用于提升边缘用户的无线链路速率。受益于 VPON 的架构，WSC 中的 RRH 通过联合传输 (JT, joint transmission) 为用户提供服务。在 JT 模式下，多个 RRH 在同一时频资源块上为同一个用户发送数据，将小区间的干扰信号转换为有用信号来提升小区边缘用户的性能^[11]。利用联合传输的优势，可将小区边缘用户请求的视频在无线域服务小区内进行分布式的协作缓存。但是，当 RRH 的负载较大时，由于缓存空间有限，使得较多用户仍需通过前传与回传获取视频，容易产生拥塞，导致用户 QoE 降低。因此，光域服务小区表示为用户协作缓存的 RRH，用于提升重负载小区用户的内容访问时延，避免重负载小区的大量用户到内容服务器获取视频。如图 1 所示，光域服务小区有 2 种传输方式，一种是通过 RRH—ONU—BBU—ONU—RRH 的转发方式来进行缓存的协作传输^[13]，如图 1 粗线箭头所示；由于本文提出架构的 ONU 与 RRH 为一对多的连接关系，另一种则可以利用 ONU 作为转发媒介，通过 RRH—ONU—RRH 的路径进行缓存的协作传输，如图 1 虚线箭头所示。

3 基于协作增益的 VPON 构建

在小区密集部署场景下，位于小区边缘的用户被多个 RRH 覆盖，受到严重的同频干扰，并且小区间的负载不均衡使得轻负载 RRH 的缓存利用率不高，而重负载 RRH 的大量用户仍需通过核心网获取视频，访问时延大。因此，为提升网络中小小区边缘用户与重负载小区用户的性能，本节提出了协作缓存与传输增益感知的 VPON 构建机制。VPON

的构建主要通过 RRH 分簇与 ONU 分簇来实现，其中，RRH 分簇实现无线域协作缓存与传输增益，ONU 分簇则实现光域协作缓存与传输增益。

3.1 RRH 分簇

对于本文所提出的 VPON 架构来说，可以采用 VPON 内的联合传输或 VPON 间的联合传输来实现 RRH 的联合传输，前者只需在同一 BBU 内部各小区间交互联合传输处理相关的业务数据和控制信息，较易于实现；而后者需要在 BBU 间交互这些信息，对 X2 接口带宽有很高要求，时延也比前者更大^[14]。因此，本文仅考虑 VPON 内的联合传输。在用户建立无线域服务小区之前，有必要对 RRH 进行分簇使得协作区域中边缘用户的无线域协作缓存与传输增益最大化。基于广度优先搜索 (BFS, breadth first search) 方法，本文提出贪婪的动态成簇算法，首先，BBU 根据边缘用户的干扰信息与小区的请求内容相似度构建连接关系图；其次，基于边权重对 RRH 进行分簇，即将连接关系图分为多个子图，每个子图代表一个联合传输簇。

1) 基于边权重的图构建

令 $h_{i,k}$ 和 $h_{j,k}$ 分别表示 RRH_i 和 RRH_j 到用户 k 的信道增益， P_k 与 P_j 分别表示有用信号流与干扰信号流， z_k 表示用户 k 的高斯白噪声，则用户 k 由本地 RRH_i 服务时的接收信号可由式(1)表示。

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

本文定义 I_{r_i,r_j} 为 RRH_i 与 RRH_j 之间的干扰程度，即 RRH_i 中的边缘用户接收 RRH_j 干扰信号的平均值，计算式如式(2)所示。RSRP_{j,k} 表示用户 k 接收到 RRH_j 的参考干扰信号接收功率， $|\mathcal{U}_i^e|$ 表示 RRH_i 接收 RRH_j 干扰信号的边缘用户数目。

$$I_{r_i,r_j} = \frac{\sum_{k \in \mathcal{U}_i^e} \text{RSRP}_{j,k}}{|\mathcal{U}_i^e|} \quad (2)$$

用 v_k^i 表示 RRH_i 中的边缘用户请求视频 k 的二进制变量，则 $N(r_i) = \{v_1^i, v_2^i, \dots, v_k^i, \dots, v_K^i\}$ 为 RRH_i 请求的视频集合，其中 K 为总视频数目。于是，RRH_i 与 RRH_j 的边缘用户请求内容相似程度 S_{r_i,r_j} 可以由杰卡德系数得出，如式(3)所示。由式(3)可知， $0 \leq S_{r_i,r_j} \leq 1$ ， S_{r_i,r_j} 数值越大，表示 2 个小区边缘用户的兴趣偏好越相似。

$$S_{r_i, r_j} = \frac{|N(r_i) \cap N(r_j)|}{|N(r_i) \cup N(r_j)|} \quad (3)$$

因此, RRH_{*i*} 到 RRH_{*j*} 的协作缓存与传输增益 w_{r_i, r_j} 可由式(4)表示, 干扰程度 I_{r_i, r_j} 表征 2 个小区间的协作传输增益, 请求内容相似度 S_{r_i, r_j} 表征协作缓存增益。

$$w_{r_i, r_j} = I_{r_i, r_j} S_{r_i, r_j} \quad (4)$$

根据协作缓存与传输增益构建图 $G = (V, E)$, RRH 表示顶点, 其权重为 RRH 的负载 (即 load_v), 2 个 RRH 之间的边权重为协作缓存与传输增益。当 $w_{r_i, r_j} = 0$ 时, 不需要构建连接。每个 RRH 具有与其他 RRH 间的权重列表 $W_i = \{w_{i,1}, \dots, w_{i,R}\}$, 由于图 G 中 2 个顶点间的有向边权重之和表示 2 个 RRH 纳入同一簇中的协作缓存与传输增益, 因此通过式(5)将有向图转换为无向图, 其中, $w_{r_i \rightarrow r_j}$ 与 $w_{r_j \rightarrow r_i}$ 为有向图的权重。

$$w_{r_i, r_j} = w_{r_i \rightarrow r_j} + w_{r_j \rightarrow r_i} \quad (5)$$

2) 基于图的 RRH 分簇

在 RRH 分簇时, 对于任意 2 个顶点 i 与 j , 若满足如式(6)所示的条件, 即 i 与 j 互为与其相邻顶点的最大边权重, 则将 2 个顶点分到同一簇中, 以实现簇内的最大协作缓存与传输增益。

$$\begin{cases} w_{r_i, r_j} = \max_{r_s} w_{r_i, r_s} \\ w_{r_i, r_j} = \max_{r_k} w_{r_k, r_j} \end{cases} \quad (6)$$

其中, s 与 k 分别表示 i 与 j 的相邻顶点。类似地, 该规则可以延伸到点到子图与子图到子图中, 则点到子图与子图到子图的权重如式(7)所示。

$$w_{i, G_1} = \sum_{k \in G_1} w_{i, k}, \quad w_{G_1, G_2} = \sum_{i \in G_1} \sum_{j \in G_2} w_{i, j} \quad (7)$$

本文提出了基于广度优先搜索的贪婪成簇算法, 如算法 1 所示。首先, 将每个顶点看作单顶点的子图, 在每次搜索阶段, 从每个子图出发, 访问该子图 (即找到其相邻子图), 并选择与该子图具有最大边权重的相邻子图作为其候选协作点。若 2 个子图满足式(6), 即候选协作点的最大边权重的相邻子图为该子图本身, 则将候选协作点与该子图放在同一簇中, 去除 2 个子图间的干扰, 并增大小区间的无线域协作缓存可能性。随后, 继续访问新子图的候选协作点。该程序重复执行, 直至子图超出波长容量。当所有子图均被

纳入簇中 (即 $\mathcal{S} = \emptyset$), 该算法终止。根据算法 1, 图将被分为多个子图, 每个子图代表一个联合传输簇, 在每个簇内能够实现最大的协作缓存与传输增益。

算法 1 RRH 贪婪成簇算法

初始化阶段

① 将每个顶点 $v \in \mathcal{V}$ 视作一个子图, 所有子图形成集合 $\mathcal{S} = \{\{v\}\}$; RRH 簇集合为 \mathcal{VP} , 且 $\mathcal{VP} = \emptyset$

搜索阶段

② while ($\mathcal{V} \neq \emptyset$ and $\mathcal{W} \neq \emptyset$)

③ 针对每个子图 $G \in \mathcal{S}$, 搜索其相邻顶点 $k \in \mathcal{V}$ 与相邻子图 $G_a \in \mathcal{S}$, 并从中找出与该子图的权重最大的顶点或子图 $\Omega(G)$

④ for each $\Omega(G)$

⑤ if $G = \Omega(\Omega(G))$ and $\text{load}_{G \cup \Omega(G)} \leq C_w$

⑥ $G = G \cup \Omega(G)$;

⑦ 将 $\Omega(G)$ 从 \mathcal{S} 集合中移除

⑧ end if

⑨ if $G = \Omega(\Omega(G))$ and $\text{load}_{G \cup \Omega(G)} > C_w$

⑩ 将 G 从 \mathcal{S} 集合中移除并添加到 \mathcal{VP}

⑪ end if

⑫ 根据式(7)更新连接子图间的边权重

⑬ end for

⑭ 移除 \mathcal{S} 中的剩余子图到簇集合 \mathcal{VP} 中

⑮ end while

3.2 ONU 分簇

针对视频业务而言, 播放中断是影响用户 QoE 的关键因素。受到 RRH 无线资源有限性的制约, 在初始缓冲阶段的最大初始缓冲时间内, 当用户实际缓冲数据小于预期缓冲数据 (即 $B_t < B_e$) 时, 将导致视频中断。在此种情况下, 为防止视频播放中断, 用户需切换视频的分辨率。显然, 视频分辨率、最大初始缓冲时间和网络负载这三者之间存在紧密联系。目前, 普遍采用基于 E-LBP 表的速率自适应算法来定义 RRH 的负载情况, 即通过小区中用户是否切换视频比特率版本来表征 RRH 的负载情况^[15]。E-LBP 表中不同的视频分辨率和传输速率对应不同的最大初始缓冲时间 D_{\max} 。本文定义 $c(u)$ 为用户 u 是否切换分辨率版本的二进制变量, 则 RRH 的轻重负载归类如式(8)所示。若在 RRH_{*r*} 中存在用户切换比特率版本, 则说明该 RRH 负载过重, 用户不得不切换分辨率版本以避免中断; 当 RRH 内所有用户在最大初始缓冲时间内未改变分辨率版

本, 则为轻负载 RRH。

$$\begin{cases} r \in \mathcal{VP}_m^h, \exists u \in \mathcal{U}_r, c(u) = 1 \\ r \in \mathcal{VP}_m^l, \forall u \in \mathcal{U}_r, c(u) = 0 \end{cases} \quad (8)$$

ONU 与 RRH 为一对多的连接关系, 不同 VPON 内的 RRH 可能连接到相同的 ONU。然而, ONU 在同一调度周期内只能调谐到一个波长, 因此需要对 ONU 进行分簇, 使得各 VPON 独立运行。ONU 的分簇必须使每个 RRH 至少被一个 ONU 所支配, 那么 ONU 的分簇可以转化为二分图的支配集问题。任一 VPON 内的所有 RRH 节点构成集合 $S = \{s_1, s_2, \dots, s_n\}$, RRH 所连接的 ONU 节点构成集合 $V = \{v_1, v_2, \dots, v_m\}$, RRH 与 ONU 之间的链路构成集合 E , 则 ONU 与 RRH 的关系可用二分图 $G = (S, E, V)$ 来表示。任一 VPON 内的 RRH 节点 s_i 由与其相连的 ONU 节点 $v_1^i, v_2^i, \dots, v_{t_i}^i$ 所支配, 其中, $v_1^i, v_2^i, \dots, v_{t_i}^i \in V$, t_i 表示支配 RRH 节点 s_i 的 ONU 数目。首先, 求解每个 VPON 的极小有效支配集, 根据式(9)可计算所有的有效支配集。

$$H = (v_1^1 \vee v_2^1 \vee \dots \vee v_{t_1}^1) \wedge (v_1^2 \vee v_2^2 \vee \dots \vee v_{t_2}^2) \wedge \dots \wedge (v_1^n \vee v_2^n \vee \dots \vee v_{t_n}^n) \quad (9)$$

式(9)表示选取支配 RRH 节点 s_i 的任意一个 ONU 节点 v_j 形成的组合为一个有效支配集。进而, 根据支配集的吸收律对式(9)进行简化, 以求解极小有效支配集。

其次, 所有 VPON 找到极小有效支配集后需分配剩余 ONU 节点。由于缓存空间有限, 轻负载 RRH 可以为重负载 RRH 的用户提供光域协作缓存空间, 增强缓存命中率。定义 ONU_n 的最大传输容量为 M_n , μ_m 表示 ONU_n 仅连接 VPON_m 中 RRH 的带宽使用率。本文采用 ONU 的剩余带宽表征光域协作传输增益, 采用轻、重负载 RRH 的平均内容请求相似度表征光域协作缓存增益。因此, 当 ONU 所连接的 RRH 在不同 VPON 时, ONU 在不同 VPON 中的权重 Ω_{mn} 为

$$\Omega_{mn} = \alpha_m \frac{\sum_{r_i \in \mathcal{R}_n^h} \sum_{r_j \in \mathcal{VP}_m^l} \frac{S_{r_i, r_j}}{|\mathcal{VP}_m^l| |\mathcal{R}_n^h|}}{M_n \mu_m} \quad (10)$$

其中, \mathcal{R}_n^h 表示 ONU_n 在 VPON_m 中所连接的重负载 RRH 集合, α_m 表示 VPON_m 中所有 ONU 的平均负载率。在具有相同光域协作增益的情况下, ONU 分

配给具有较大平均负载率的 VPON, 实现 ONU 的负载均衡。 ONU_n 在 VPON_m 中的增益与 ONU 的使用带宽量成反比, ONU 负载越高, 排队时延越大, 光域协作传输增益越小; ONU_n 在 VPON_m 中的增益与轻、重负载 RRH 的平均内容请求相似度成正比。ONU 被纳入增益最大的 VPON 中, 然后通过关闭 ONU 的接口断开该 ONU 与其他 VPON 中 RRH 的固有连接, 从而形成独立的 VPON。

4 联合前传带宽与缓存分配机制

QoE 是指用户对应用或业务的质量和性能的综合主观感受, 反映用户对业务的可接受程度。初始访问时延直接影响用户对该视频的满意程度, 若访问时延过长, 用户可能放弃对视频的播放。显然, 静态缓存配置方式无法满足用户的时变需求, 本文提出联合前传带宽与缓存分配机制, 以达到降低下一调度周期内用户访问时延的目的。为此, 首先, 通过考虑用户行为与光域的协作缓存计算视频的缓存价值; 其次, 根据 ONU-RRH 的连接状态定义请求视频的离散度, 用于评估预缓存过程的带宽利用率; 最后, 基于视频的缓存价值成本比通过多次迭代完成视频的预缓存。

4.1 视频缓存价值分析

视频缓存价值与用户行为紧密相关, 因此定义 χ_{ir} 为 RRH_r 中的缓存内容被用户 i 所命中的概率, 该参数取决于用户 i 的移动模式 (即当前位置、移动速度和移动方向) 及请求视频的流行度。当给定用户请求视频的流行度时, 其移动速度越快或位于小区边缘, 由 RRH_r 所服务的概率越小; 当给定用户的移动模式时, 请求视频的流行度越高, 使用该 RRH 缓存视频的概率越大。 χ_{ir} 能够通过机器学习对用户行为进行预测, 因此用户请求缓存的概率集合 $\{\chi_{ir}\}$ 对于系统来说是已知的。由于大多数的视频请求集中在少数的视频内容, 因此用户视频请求服从参数 $\tau = 0.56$ 的 Zipf 分布, 则用户请求视频 v 的概率为

$$f_v = \frac{1}{v^\tau} \frac{1}{\sum_{m=1}^v \frac{1}{m^\tau}} \quad (11)$$

其中, $\tau = 0$ 意味着视频流行度服从均匀分布, τ 越大, 视频请求集中在更少的流行视频上。根据文献[16], 可利用回声状态网络模型预测调度周期内用户 i 请求视频 v 的概率 p_{iv} 以及用户 i 由 RRH_r 服务的概率

s_{ir} , 因此 RRH_r 中的缓存内容被用户 i 所命中的概率 χ_{ir} 可由式(12)计算, $\chi_{ir} \in (0,1)$ 。

$$\chi_{ir} = f_v p_{iv} s_{ir} \quad (12)$$

轻负载 RRH 能为重负载 RRH 提供光域协作缓存, 为避免重负载 RRH 过载, 定义重负载 RRH_r 接收其他轻负载 RRH 的缓存带宽为 B_r 。假设视频 v 的大小为 s_v , 在调度周期内所需的传输带宽为 b_v , RRH_r 的缓存空间满足以下 2 种条件时需对视频 v 进行预缓存。

1) 直接缓存条件: 缓存剩余空间大小 $C_r' > 0$, 且 $C_r' \geq s_v$ 。

2) 缓存替换条件: 不满足条件 1), 但存在已缓存视频 v' 的请求概率比视频 v 的请求概率小, 且满足 $s_{v'} > s_v$; 或者存在 n 个视频的请求概率之和比视频 v 的请求概率小, 且满足 $s_1 + \dots + s_n > s_v$ 。

针对重负载 RRH_r , 若视频 v 满足上述条件, 则进行缓存, 其缓存价值定义为 g_{rv}^h , 计算式为

$$g_{rv}^h = \frac{gl_{rv}}{s_v} = \frac{\sum_{i \in \mathcal{I}_r} a_{iv} \chi_{ir}}{s_v} \quad (13)$$

其中, a_{iv} 为二进制变量, 表征用户 i 是否请求视频 v ; \mathcal{I}_r 为 RRH_r 中的用户集合。缓存价值与 RRH_r 内用户请求缓存视频 v 的概率和 gl_{rv} 成正比, 与视频 v 所占的缓存空间大小 s_v 成反比。

在 $VPON$ 内, 假设重负载 RRH 与轻负载 RRH 的数目分别为 M_1 、 M_2 。针对 $VPON$ 中具有光域协作缓存可能性的重负载 RRH , 其请求缓存视频 v 的概率之和为 gc_v , 计算式为

$$gc_v = \sum_{r=1}^{M_1} (1 - b_{rv}) \sum_{i \in \mathcal{I}_r} a_{iv} \chi_{ir} [B_r' \geq b_v] \quad (14)$$

其中, b_{rv} 为二进制变量, 表示视频 v 是否满足在 RRH_r 缓存的条件; 当重负载 RRH_r 的剩余接收带宽 B_r' 大于视频 v 的传输带宽时, $[B_r' \geq b_v] = 1$, 否则 $[B_r' \geq b_v] = 0$ 。针对具有光域协作缓存的重负载 RRH , 视频 v 选择具有最大缓存价值的轻负载 RRH_r 进行协作缓存, 则重负载 RRH 中请求视频 v 的用户其光域服务小区为 RRH_r 。由此, 可通过式(15)计算轻负载 RRH_r 的缓存价值。

$$g_{rv}^l = \frac{gl_{rv} + gc_v}{s_v} \quad (15)$$

根据上述分析, 在 $VPON_m$ 内视频 v 的缓存价

值可通过式(16)进行计算。

$$gp_v = \sum_{r \in VP_m^h} b_{rv} g_{rv}^h + \sum_{r \in VP_m^l} b_{rv} g_{rv}^l \quad (16)$$

4.2 请求视频离散度

光域 ONU 与无线域 RRH 为一对多的连接关系, 同理 RRH 也可以与多个 ONU 相连。对于多播视频业务, 若多个 RRH 同时申请缓存相同的视频, 则在网络中会生成相应的多播路由表, 不同的视频业务对应不同的多播路由表。请求视频 v 的 RRH 越分散, 即需要的 ONU 支配点越多, 越容易造成数据的冗余传输, 浪费 ONU 的带宽资源。因此, 本文根据 ONU - RRH 的连接状态定义了请求视频的离散度 MD_v , 反映缓存该视频时 ONU 传输带宽的利用率, 计算式如式(17)所示。

$$MD_v = \frac{N_v}{N_m} \quad (17)$$

其中, N_v 表示支配请求缓存视频 v 的 RRH 所需 ONU 的最小支配数目, N_m 表示 $VPON_m$ 中 ONU 的总数目, 显然, $0 < MD_v \leq 1$ 。视频 v 的离散度越小, 请求视频 v 的 RRH 集中度越高, 多播性能越好。通过 ONU 的多播能够有效提高带宽利用率, 同时减少数据的冗余传输。

4.3 联合前传带宽与缓存分配算法

动态预缓存过程不仅与视频的缓存价值有关, 还要考虑预缓存的带宽成本, 包括下行波长带宽与 ONU 的传输带宽。因此, 定义缓存视频 v 的价值成本比 ϕ_v , 如式(18)所示。显然, 价值成本比与视频的缓存价值成正比, 而与视频所需的传输带宽以及视频的离散度成反比。

$$\phi_v = \frac{gp_v}{MD_v b_v} \quad (18)$$

4.1 节中缓存价值并未考虑前传带宽与 ONU 传输带宽的可用性, 换言之, 当前传带宽或 ONU 的剩余带宽小于视频 v 的传输带宽时, RRH_r 不具备缓存视频 v 的价值, 因此对式(16)进行改写, 即

$$gp_v = \sum_{r \in VP_m^h} b_{rv} g_{rv}^h + \sum_{r \in VP_m^l} b_{rv} g_{rv}^l \quad (19)$$

$$\text{s.t. } C_w' \geq b_v, \exists c_{or} B_o' \geq b_v$$

其中, C_w' 表示剩余波长容量, B_o' 表示 ONU_o 的剩余传输带宽, c_{or} 表示 ONU_o 与 RRH_r 的连接状态二进制变量。根据式(19)可计算视频 v 的价值成本比, 则联合前传带宽与缓存分配算法可根据各视频的价值成

本比的降序进行预缓存。该算法优先选择具有最高价值成本比的视频 v 进行缓存, 其中求取的最小支配数目 N_v 必须满足 ONU 的传输带宽约束。其次, 根据重负载 RRH 的剩余接收带宽 B'_r 、剩余前传带宽 C'_w 与剩余 ONU 的传输带宽 B'_o 分别更新 g_{c_v} 、 g_{r_v} 与 MD_v , 并重新计算 Φ_v 。最后, 多次迭代对视频进行预缓存, 直至不需要更新缓存或带宽已耗尽为止。

5 QoE 感知的无线带宽分配机制

无线带宽的分配不仅影响用户的初始访问时延, 还直接决定用户对视频质量的感知, 因此本节提出 QoE 感知的无线带宽分配机制。根据第 4 节所获知的缓存状态信息, 首先, 为边缘用户建立无线域服务小区将 RRH 与用户相关联, 从而降低用户的访问时延并保障较高信噪比; 其次, 每个用户的 QoE 优化是独立且自私的, 这就使得无线资源利用率低、QoE 较差, 因此本文提出缓冲水平感知的无线带宽分配, 实现用户间无线资源的灵活配置, 进一步提高 QoE。

5.1 用户无线域服务小区的建立

为提升用户的 QoE, 无线域服务小区结合 RRH 的缓存状态为边缘用户提供 2 种传输方式。如图 2(a) 所示, 当无线域服务小区内的所有 RRH 均缓存用户请求的内容, 则所有 RRH 通过联合传输为用户提供视频服务; 如图 2(b) 所示, 当无线域服务小区内部分 RRH 缓存了用户请求的视频时, 则利用已缓存的部分 RRH 为用户联合传输, 并抑制未缓存该视频的 RRH 使用该视频资源块, 消除干扰, 该方式不仅提升了用户的无线数据率, 还避免了未缓存的 RRH 到内容服务器获取该视频所产生的时延, 减小了用户的访问时延。

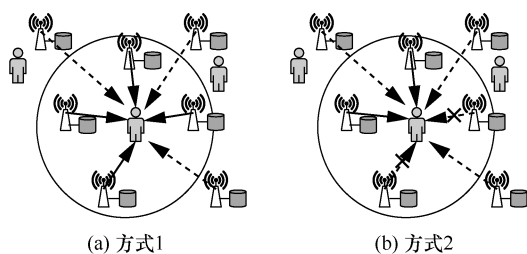


图 2 无线域服务小区协作传输方式

根据无线域服务小区的 2 种传输方式, 可通过该小区的平均信噪比 $\overline{\text{SINR}}_r$ 为边缘用户建立与 RRH 的关联。假设边缘用户 i 的参考信号接收功率集合为 RSRP_i , 并以降序进行排列。通过按序纳入 RSRP_i 中对应的 RRH 为用户 i 建立无线域服务小区

WSC_i , 直至 $\text{SINR}_i \geq \overline{\text{SINR}}_r$, 其信噪比的计算如式(20)所示。

$$\text{SINR}_i = \frac{\sum_{r \in \text{WSC}_i} c_{ir} g_{ir} p_r}{\sigma_0 + \sum_{m \in \text{WSC}_i} g_{im} p_m} \quad (20)$$

其中, 当用户 i 请求的视频已在 RRH_r 缓存时, $c_{ir} = 1$, 否则 $c_{ir} = 0$; σ_0 表示高斯白噪声的噪声功率; g_{ir} 表示信道增益; p_r 表示 RRH_r 的发射功率。通过为边缘用户建立无线域服务小区, 从而确定每个 RRH 与用户的关联集合 $\mathcal{U}_r^{\text{wir}}$ 。

5.2 缓冲水平感知的质量自适应

RRH 无线带宽的分配直接决定关联用户对视频质量的感知, 本文采用平均意见得分 (MOS, mean opinion score) 表示实时视频流的用户满意度。视频流的效用函数 Q 定义为发送数据率 γ 的函数, 如式(21)所示。

$$Q = f(\gamma), f: \gamma \rightarrow \text{MOS} \quad (21)$$

根据文献[17]的测试报告, 峰值信噪比 (PSNR, peak signal to noise ratio) 与 MOS 之间可视为简单的线性映射关系。因此, MOS 值可取 1.0 (即 PSNR 为 30 dB) ~ 4.5 (即 PSNR 为 42 dB) 之间的任意值, 1.0 表示最差的用户满意度, 4.5 表示最佳的用户满意度。由于不同的视频序列具有不同的图像特征, 因此效用函数 Q 采用文献[18]的参数模型, 利用 PSNR 衡量视频质量, 并通过有限的经验数据 (即 3 对峰值信噪比与发送数据率值) 推导模型中的参数值, 从而确定 PSNR 与 γ 的函数关系。本文的目标是确定优化周期中每个用户的无线资源份额 δ_i , 即物理资源块的占比。假设用户端无差错地接收到所有数据, 于是对于任意用户 $i \in \mathcal{U}_r^{\text{wir}}$ 的无线数据率如式(22)所示, 其中, B_{\max} 为 RRH 的最大可用带宽。

$$\gamma_i = g(\delta_i) = \delta_i B_{\max} (1 + \text{SINR}_i) \quad (22)$$

QoE 感知的无线资源分配策略的目标是确定与 RRH 所关联用户的传输速率, 从而最大化用户的满意度。因此, 优化问题如式(23)所示, 该优化问题可采用贪婪搜索算法进行求解, δ_{opt} 表示最优的资源分配组。

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

其中, N 为每个 RRH 关联的用户数目 ($N = |\mathcal{U}_r^{wir}|$)。 $g(\delta_i) \geq R_{\min,i}$ 表示无线带宽的分配必须满足用户最基础的视频质量, 若 $g(\delta_i) < R_{\min,i}$, 则用户无法解码最低分辨率版本, 同时造成无线资源的浪费。 P_i 为用户 i 的惩罚项, 表示用户 i 对质量波动的感知, 如式(24)所示。用户在连续调度周期的较大质量波动会造成视频的不平滑播放, 影响用户的 QoE。如果 2 个连续的优化周期 t 与 $t-1$ 之间的质量变化超过最小可觉差 (JND, just noticeable difference) 阈值 λ_{th} 时, 则会增加一个质量波动的惩罚项 P_i 。根据文献[17]的测试结果, 取阈值 $\lambda_{th} = 0.23$ MOS。当 $|Q_i(g(\delta_i))_t - Q_i(g(\delta_i))_{t-1}| \leq \lambda_{th}$ 时, 质量的波动在用户不可感知的变化范围内。

$$P_i = \max\{0, |Q_i(g(\delta_i))_t - Q_i(g(\delta_i))_{t-1}| - \lambda_{th}\} \quad (24)$$

根据式(23)的无线速率分配为用户选择播放分辨率版本 φ_i , 使 $\varphi_i < \gamma_i$, 从而保证视频播放不中断。该方式下用户的 QoE 并非最佳, 原因是用户设备的缓冲区具有一定的缓存, 根据其缓冲水平 (BL, buffer level) 可以选择更高的分辨率版本。缓冲水平定义为用户端缓冲区已缓冲的视频时间长度, 由此可确定用户能够选择的分辨率版本范围, 如式(25)所示。

$$\gamma_{\min,i} \leq \varphi_i \leq \gamma_i \left(1 + \frac{BL_i}{T}\right) \quad (25)$$

式(25)表示在优化周期 T 内, 为了保证视频播放不中断, 输入数据量必须大于或等于用户的消耗数据量, 即 $\varphi_i T \geq \gamma_i (BL_i + T)$ 。因此, 可根据式(23)所获取的资源分配 γ_i 以及用户的缓冲水平 BL_i 确定用户 i 的最高分辨率版本, 从而最大化用户满意度。

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

显然, 缓冲水平感知的分辨率版本选择有利于提高用户的满意度, 根据式(26)选择最高比特率版本虽然增大了用户满意度, 但无线信道的动态变化不利于缓冲水平较低的用户。可见, 缓冲水平的构建能够补偿无线信道的动态变化。可实现 QoE 表示在不考虑缓冲水平的情况下, 用户可实现的 MOS 值。可实现 QoE 较高但缓冲水平较低的用户可以切换到较低的比特率版本, 填充缓冲区; 相反, 可实

现 QoE 较低但具有较高缓冲水平的用户可以切换到高于传输速率的分辨率版本。因此, 本文定义了与缓冲水平相关联的 QoE 上下限, 分别用 UB 与 LB 表示。为简化问题复杂度, UB、LB 与缓冲水平 BL_i 为线性正比关系, 即

$$\begin{cases} LB = Q_L + a BL_i \\ UB = Q_H + b BL_i \end{cases} \quad (27)$$

其中, Q_L 与 Q_H 分别取值 1 与 4, a 与 b 分别表示用户消耗缓冲的速度与建立缓冲的速度。当 $Q_L = 1$ 时, 对应的 LB 表示在最低传输速率的情况下, 随着缓冲水平的增大, 用户能够实现的最小 MOS 值; $Q_H = 4$ 保证无线资源有利但缓冲水平较低的用户的质量不会有太大的降低。根据式(23)的传输速率 γ_i , 得到用户可实现的 MOS 值为 $\mu_i = Q_i(\varphi_i)$, 其中 $\varphi_i \leq \gamma_i$ 。于是, 根据 μ_i 的取值范围定义缓冲水平阈值 κ_{th} 。

1) 当 $Q_L \leq \mu_i \leq Q_H$, 即用户处于较低 QoE 时, 缓冲水平阈值 κ_{th} 计算如式(28)所示, 其中, a 决定用户 i 消耗缓冲的速度。

$$\kappa_{th} = \frac{\mu_i - Q_L}{a} \quad (28)$$

式(28)表明用户的传输速率或可实现的 MOS 值越高, 缓冲水平阈值越大; 消耗缓冲的速度越快, 缓冲水平阈值越小。因此, 用户选择分辨率版本可通过缓冲阈值来进行调控, 如式(29)所示。当缓冲水平大于阈值时, 表明有足够的缓冲支持用户观看更高的比特率版本 φ_i ; 当缓冲水平低于阈值时, 保持高于可实现传输速率的最低分辨率版本即可。

$$\begin{cases} Q_i(\varphi_i) \geq Q_L + a BL_i, BL_i \geq \kappa_{th} \\ Q_i(\varphi_i) \geq \mu_i, BL_i < \kappa_{th} \end{cases} \quad (29)$$

2) 当 $\mu_i > Q_H$, 即用户处于较高 QoE 时, 缓冲水平阈值 κ_{th} 计算如式(30)所示, 其中, b 决定用户 i 建立缓冲的速度。

$$\kappa_{th} = \frac{\mu_i - Q_H}{b} \quad (30)$$

式(30)表明用户的传输速率或可实现的 MOS 值越高, 缓冲水平阈值越大; 建立缓冲的速度越快, 缓冲水平阈值越小。当缓冲水平小于阈值时, 表明缓冲水平低, 难以支持用户观看更高的比特率版本 φ_i ; 当缓冲水平高于阈值时, 保持低于可实现传输速率的最低分辨率版本即可。

$$\begin{cases} Q_i(\varphi_i) \leq \mu_i, BL_i \geq \kappa_{th} \\ Q_i(\varphi_i) \leq Q_H + bBL_i, BL_i < \kappa_{th} \end{cases} \quad (31)$$

5.3 缓冲水平感知的联合带宽分配与质量自适应

根据用户缓冲水平灵活地重配置无线带宽能够进一步提高用户满意度，例如在不降低该用户请求比特率版本的情况下，可将缓冲水平高的用户的部分带宽资源给予较高分辨率流且缓冲水平低的另一用户。与 5.2 节分别确定传输速率与播放速率不同，优化问题(32)根据缓冲水平联合配置传输速率与播放速率来提升用户的 QoE。

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

本文提出贪婪算法求解问题(32)，通过多次迭代得到最优的无线资源配置策略。首先分别根据式(23)与式(28)~式(31)初始化无线带宽配置与分辨率选择集合，即 $\{\gamma_1, \dots, \gamma_N\}^0$ 、 $\{\varphi_1, \dots, \varphi_N\}^0$ ，并且定义用户的效用函数为 $U_i = Q_i(\gamma_i, BL_i) - P_i$ 。紧接着，在每次迭代过程中，搜索用户 i 与用户 j ，使得用户 i 与用户 j 在分别增加与减少无线份额 $\Delta\delta$ 时，2 个用户的效用增量与效用减量之差最大化，如式(33)所示。

$$\arg \max_{i,j} \Delta U_{ij} = \Delta U_i(g_i(\delta_i + \Delta\delta)) - \Delta U_j(g_j(\delta_j - \Delta\delta)) \quad (33)$$

确定用户对后，分别更新 $(\gamma_i)^m$ 、 $(\varphi_i)^m$ 与 $(\gamma_j)^m$ 、 $(\varphi_j)^m$ 。当用户对的效用增量 ΔU_{ij} 极小时（即 $\Delta U_{ij} \leq \Delta U_{min}$ ），停止迭代，输出 $\{\delta_1, \dots, \delta_N\}$ 。边缘用户由多个 RRH 提供无线资源，因此，边缘用户 k 取其无线域服务小区中 RRH 的最小无线带宽份额，即 $\delta_i = \min\{\delta_{ir}, r \in WSC_k\}$ 。此时，每个 RRH 会剩余带宽资源，于是固定边缘用户无线带宽份额，根据上述贪婪算法对资源进行重分配。

6 数值分析

本文使用 NS2 仿真平台对提出的 QoE 感知的视频协作缓存与传输机制（QACCTM, QoE-aware video cooperative caching and transmission mechanism）进行验证，并与文献[11]提出的 CEC 算法、文献[19]提出的 QAWEC 算法以及文献[15]提出的 RARCP 算法做对比。CEC 算法根据静态缓存信息

为边缘用户选取合适的 RRH 集合并为其提供联合传输服务，保证用户在无线链路速率的情况下最小化回传流量。QAWEC 算法根据用户行为建立带宽配置与边缘缓存的联合优化问题，以保证缓存命中率与用户体验质量。RARCP 算法制定用户偏好（UPP, user preference profile）感知的主动预缓存策略，进而基于缓存信息调度回传与无线资源，从而最大化用户 QoE。仿真参数设置如表 1 所示。

表 1 仿真参数设置

参数设定	参数数值
RRH-ONU 数量	20
RRH 覆盖范围/m ²	500 × 500
视频数目	10 ⁴
波长容量/(Gbit · s ⁻¹)	10
平均无线接入时延/ms	50
平均前传时延与平均回传时延/ms	20, 80
路径损耗/dB	$L=128+37.6lgR$
传输功率/dBm	20
RRH 缓存容量/GB	100
系统带宽/MHz	20
视频分辨率版本	11

6.1 不同用户数量下的性能分析

边缘用户吞吐量是衡量用户 QoE 的关键因素之一，定义为单位时间内无线域服务小区为边缘用户联合传输时成功传送的数据量。图 3 为不同用户数目与边缘用户平均吞吐量的关系。由图 3 可以看出，随着用户数目的增多，边缘用户的平均吞吐量呈下降趋势。本文提出的 QACCTM 算法比其他算法性能提高 30%左右，这主要是因为，在 VPON 构建时考虑了无线域的协作缓存与传输增益，使得边缘用户能够利用无线域服务小区中的缓存内容为其提供联合传输服务，并且不需要从内容服务器中去获取缓存。这不仅增强了边缘用户的数据链路速率，还避免了从内容服务器传输视频所产生的分组丢失率，从而增大了边缘用户的吞吐量。然而，CEC 算法所构建的联合传输簇使得部分 RRH 仍需通过内容服务器获取缓存，因而其性能略低于 QACCTM 算法。QAWEC 算法与 RCRCP 算法由于未考虑边缘用户的性能特征，因此吞吐量较差。

图 4 反映了不同用户数目与用户平均初始访问时延的关系。由图 4 可以看出，随着用户数目的增多，用户的平均初始访问时延呈上升趋势。但是，CEC 算法由于缓存内容的不变性，使得用户数量的

变化对用户平均访问时延的影响不大，资源利用率低。RARCP 算法通过用户内容偏好动态低更新缓存，忽略了用户的移动模式对于缓存命中率的重要性。QAWEC 算法没有关注轻、重负载小区的缓存命中率的差异，仅针对单个小区缓存内容进行优化，未有效利用轻负载小区的缓存资源。然而，QACCTM 算法综合考虑了用户的行为特征（请求偏好与移动模式）及无线域与光域的协作缓存与传输，通过构建光域或无线域服务小区降低了重负载小区用户以及边缘用户的访问时延。因此，本文提出的 QACCTM 算法相比于 RARCP 算法和 QAWEC 算法访问时延分别降低了 27.1% 和 15.9% 左右。

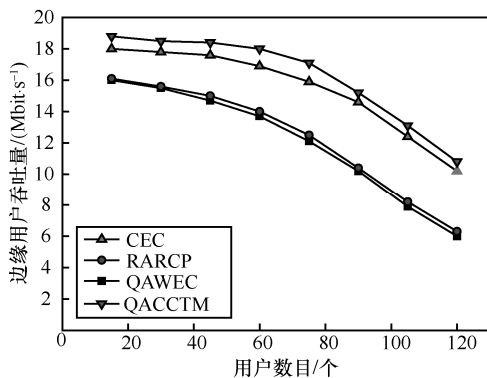


图 3 不同用户数目与边缘用户平均吞吐量的关系

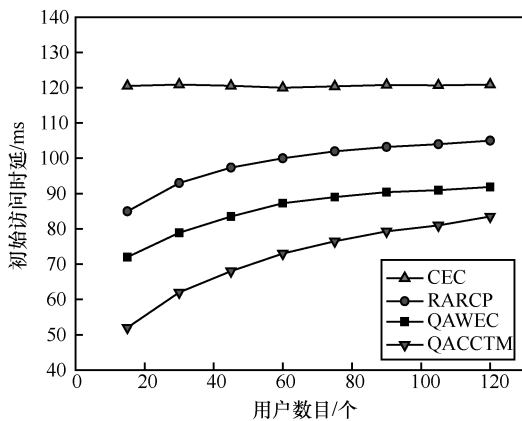


图 4 不同用户数目与用户平均初始访问时延的关系

6.2 不同缓存容量下的性能分析

缓存容量是影响用户访问时延的重要因素。当网络中所有用户请求缓存内容的概率 $\chi_{ir} \in (0.5, 0.99]$ 时，用户行为为较易获取；当 $\chi_{ir} \in (0, 0.5]$ 时，用户行为难以捕获。图 5 表示在不同用户行为下缓存容量与平均初始访问时延的关系。由图 5 可知，在 $\chi_{ir} \in (0.5, 0.99]$ 的场景下，用户的平均访问时延更短，因为用户的行为越易获取，动态预缓存的命中

率越高。由于 CEC 算法为静态缓存模式，随着缓存容量的增加，其访问时延有略微下降的趋势。然而，QACCTM 算法的访问时延对缓存容量并不敏感，原因是根据用户的行为特征进行动态预缓存极大地提高了缓存空间的利用率。

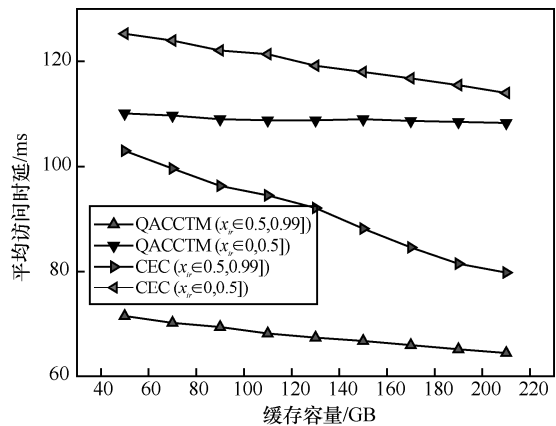


图 5 在不同用户行为下缓存容量与平均初始访问时延的关系

图 6 反映了在不同算法下缓存容量与平均初始访问时延之间的关系。由图 6 可知，RARCP 算法的访问时延相对较高，原因是在 $\chi_{ir} \in (0, 0.5]$ 的场景下，用户行为难以捕获，该算法仅考虑用户请求内容偏好，而未考虑用户的移动模式，缓存命中率低。与 QAWEC 算法和 QACCTM 算法相比，RARCP 算法对缓存容量更敏感，因为缓存容量的提升使得在该场景下用户的缓存命中率更高。然而，QAWEC 算法与 QACCTM 算法对缓存容量并不敏感，原因是 2 种算法均综合考虑了用户的行为特征进行预缓存，提高了缓存命中率，因此，即使缓存容量增加，对所增加的缓存空间的利用率并不高。此外，本文提出的算法通过光域与无线域协作缓存的方式潜在地增加了每个 RRH 的缓存容量，即使在缓存容量较低的情况下，也能实现较优的性能。

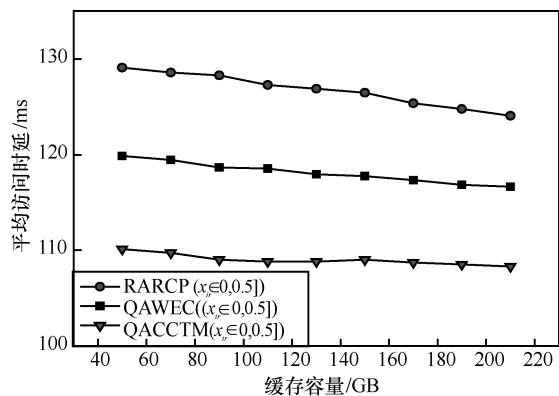


图 6 在不同算法下缓存容量与平均初始访问时延之间的关系

6.3 用户 QoE 的性能分析

视频质量是评估用户 QoE 的关键因素, 图 7 为不同用户数目与用户平均 MOS 值的关系。由于 CEC 算法未考虑无线资源分配对用户观看视频体验的影响, 因而本文仅对其他 3 种算法做对比分析。由图 7 可以看出, 随着用户数目的增多, 用户平均 MOS 值呈下降趋势。QAWEC 算法与 RCRAP 算法的性能较差, 原因是随着用户规模的增大, 为了避免视频中断, 2 种算法均会大大地降低用户观看的视频分辨率版本。然而, 本文提出的 QACCTM 算法的性能较其他算法提高了 5.6%, 该算法通过缓冲水平灵活地配置无线带宽资源, 在保证视频连续播放性的同时, 使得缓冲水平高的用户能够请求更高的比特率版本, 缓冲水平低的用户可调用缓冲水平较高用户的带宽资源用于填充缓冲区, 并且维持用户当前播放的分辨率版本。

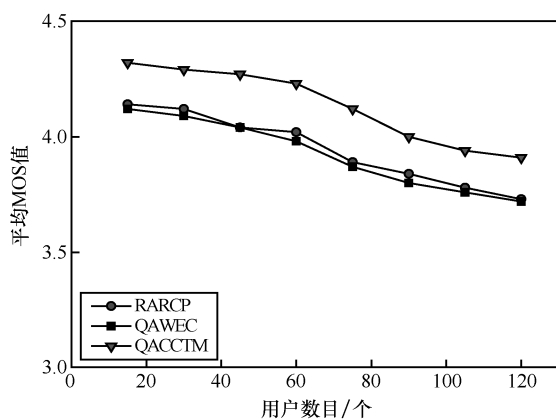


图 7 不同用户数目与用户平均 MOS 值的关系

视频的播放平滑性也是衡量用户 QoE 的重要参数。如果相邻调度周期内视频分辨率版本相差太大, 则会造成视频的非平滑播放。在视频播放期间, 非平滑切换的次数越多, 视频的平滑性越差, 这会降低用户 QoE。本文将超过一个分辨率等级的质量切换视为非平滑切换的度量, 图 8 表示不同算法下用户非平滑切换次数的累计分布函数 (CDF, cumulative distribution function)。由图 8 可知, RARCP 与 QAWEC 算法用户的非平滑切换较频繁, 因为视频的分辨率版本随瞬时传输速率的变化而变化, 其中 RARCP 算法略优于 QAWEC 算法, 原因是 RARCP 算法能够通过缓冲区的阈值指针来监控无线信道的动态变化, 从而判别是否需要切换分辨率版本, 减少非平滑切换次数。本文提出的 QACCTM 算法具有最优的性能, 因为根据用户的缓冲水平可

以灵活地配置无线资源并为用户建立缓冲区, 从而进一步减轻了无线链路速率对质量切换的影响。

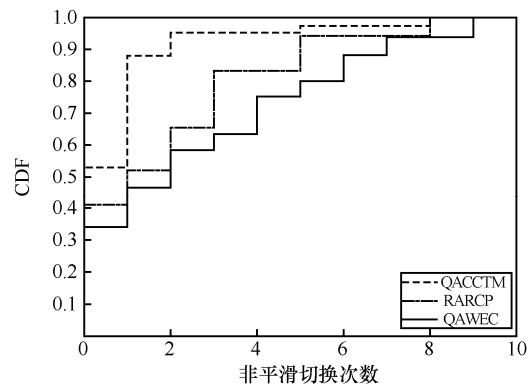


图 8 不同算法下用户非平滑切换次数的累计分布函数

7 结束语

为充分利用网络资源, 提高多媒体业务的缓存与传输性能, 在基于 TWDM-PON 与 C-RAN 的联合架构下, 本文提出了一种 QoE 感知的视频协作缓存与传输机制。首先利用小区间干扰程度、请求内容相似度、ONU 的传输带宽表征光域与无线域协作缓存与传输增益, 通过 RRH 与 ONU 分簇形成协作感知的虚拟无源光网络。进而, 联合优化用户的视频体验、带宽配置与缓存策略。在光侧, 提出了联合前传带宽与缓存分配机制, 通过动态预缓存降低用户的访问时延; 在无线侧, 根据缓冲水平实现了无线带宽的灵活配置, 增强了用户观看视频的质量。仿真结果表明, 所提机制在增强用户体验质量的同时, 有效地提高了缓存命中率。

参考文献:

- [1] CHIH-LIN I, HUANG J R, DUAN R, et al. Recent progress on C-RAN centralization and cloudification[J]. IEEE Access, 2014, 2: 1030-1039.
- [2] WU J, ZHANG Z, HONG Y, et al. Cloud radio access network (C-RAN): a primer[J]. IEEE Network, 2015, 29(1): 35-41.
- [3] HU W, YI L, HE H, et al. Soft-stacked PON for soft cloud-RAN[J]. IEEE/OSA Journal of Optical Communications & Networking, 2016, 8(11): 12-20.
- [4] HE S, TIAN H, LYU X, et al. Distributed cache placement and user association in multicast-aided heterogeneous networks[J]. IEEE Access, 2017, 5: 25365-25376.
- [5] WANG X, CHEN M, TALEB T, et al. Cache in the air: exploiting content caching and delivery techniques for 5G systems[J]. IEEE Communications Magazine, 2014, 52(2): 131-139.
- [6] ARGYRIOUS A, POULARAKIS K, IOSIFIDIS G, et al. Video delivery in dense 5G cellular networks[J]. IEEE Network, 2017, 31(4): 28-34.

- [7] AHLEHAGH H, DEY S. Video caching in radio access network: impact on delay and capacity[C]//IEEE Wireless Communications and Networking Conference. IEEE, 2012: 2276-2281.
- [8] ZHANG Z, LIU D, YUAN Y. Layered hierarchical caching for SVC-based HTTP adaptive streaming over C-RAN[C]//IEEE Wireless Communications and Networking Conference. IEEE, 2017: 1-6.
- [9] LI X, KANONAKIS K, CVIJETIC N, et al. Joint bandwidth provisioning and cache management for video distribution in software-defined passive optical networks[C]//IEEE Optical Fiber Communications Conference and Exhibition. IEEE, 2014: 1-3.
- [10] RAHMAN M, YANIKOMEROGLU H. Enhancing cell-edge performance: a downlink dynamic interference avoidance scheme with inter-cell coordination[J]. IEEE Transactions on Wireless Communications, 2010, 9(4): 1414-1425.
- [11] YU Y J, TSAI W C, PANG A C. Backhaul traffic minimization under cache-enabled CoMP transmissions over 5G cellular systems[C]//IEEE Global Communications Conference. IEEE, 2017: 1-7.
- [12] LI Z, BEGEN A C, GAHM J, et al. Streaming video over HTTP with consistent quality[C]//IEEE Association for Computing Machinery. IEEE, 2014: 248-258.
- [13] TRAN T X, HAJISAMI A, POMPILI D. Cooperative hierarchical caching in 5G cloud radio access networks (C-RAN)[J]. IEEE Network, 2017, 31(4): 35-41.
- [14] WANG X, CAVDAR C, WANG L, et al. Joint allocation of radio and optical resources in virtualized Cloud RAN with CoMP[C]//IEEE Global Communications Conference. IEEE, 2017: 1-6.
- [15] PEDERSEN H A, DEY S. Enhancing mobile video capacity and quality using rate adaptation, RAN caching and processing[J]. IEEE/ACM Transactions on Networking, 2016, 24(2): 996-1010.
- [16] CHEN M, SAAD W, YIN C, et al. Echo state networks for proactive caching in cloud-based radio access networks with mobile users[J]. IEEE Transactions on Wireless Communications, 2016, 16(6): 3520-3535.
- [17] THAKOLSRI S, KELLERER W, STEINBACH E. QoE-based cross-layer optimization of wireless video with unperceivable temporal video quality fluctuation[C]//IEEE International Conference on Communication. 2011: 1-6.
- [18] CHOI L U, IVRLAC M T, STEINBACH E, et al. Sequence-level models for distortion-rate behavior of compressed video[C]//IEEE International Conference on Image Processing. IEEE, 2005: II-486-9.
- [19] LIANG C, HE Y, YU R, et al. Enhancing QoE-aware wireless edge caching with software-defined wireless networks[J]. IEEE Transactions on Wireless Communications, 2017, 16(10): 6912-6915.

[作者简介]



吴大鹏（1979—），男，黑龙江大庆人，博士，重庆邮电大学教授、博士生导师，主要研究方向为泛在网络、互联网服务质量控制等。



李雪（1993—），女，重庆人，重庆邮电大学硕士生，主要研究方向为光无线融合网络。



李红霞（1969—），女，重庆人，中国联合网络通信有限公司重庆市分公司高级工程师，主要研究方向为光无线融合网络。