

基于多 MEC 协作的移动 VR 视频缓存和传输网络架构



郑成渝¹, 夏靖雯², 陈路遥², 唐勇^{2*}, 杨挺², 刘强³

(1. 中国电信集团公司四川分公司 成都 610000;

2. 电子科技大学计算机科学与工程学院 成都 611731;

3. 澳门科技大学资讯科技学院 中国 澳门 999078)

【摘要】针对移动虚拟现实 (VR) 视频在传统传输架构下分发时延过大问题, 提出一种基于多 MEC 协作缓存和传输的网络架构。首先, 考虑到移动 VR 视频在观看时需要计算处理而移动 VR 设备并不一定具备相应的处理能力, 提出让传输架构中的 MEC 来承担这部分计算任务。然后, 建立以最小化平均分发时延为优化目标的混合整数线性规划缓存模型, 通过商业求解器 Gurobi 来获得移动 VR 视频的最佳缓存位置。最后, 通过数值仿真实验证明, 与传统的缓存算法相比较, 提出的缓存和传输策略能有效降低移动 VR 视频的平均分发时延。

关键词 360° 视频; 内容分发; 移动边缘计算; 虚拟现实

中图分类号 TN919.8 文献标志码 A doi:10.12178/1001-0548.2022199

Mobile VR Video Caching and Streaming Network Architecture Based on Multi-MEC Collaboration

ZHENG Chengyu¹, XIA Jingwen², CHEN Luyao², TANG Yong^{2*}, YANG Ting², and LIU Qiang³

(1. China Telecom Sichuan Branch Chengdu 610000;

2. School of Computer Science and Engineering, University of Electronic Science and Technology of China Chengdu 611731;

3. Faculty of Innovation Engineering, Macau University of Science and Technology Macau China 999078)

Abstract In this paper, a network architecture based on multi-MEC (mobile edge computing) cooperative caching and streaming is proposed to address the problem that the delivery latency of mobile virtual reality (VR) video is too high under traditional delivery architecture. First of all, considering that the mobile VR video needs computing processing when watching and the mobile VR device may not always have the corresponding processing capacity, the MEC in the transmission architecture undertakes the needed computing task. Then, a mixed-integer linear programming (MILP) cache model with the optimization goal of minimizing the average delivery latency is established, and the optimal cache placement for mobile VR videos is obtained through the commercial solver Gurobi. Finally, numerical simulation experiments show that the proposed caching and streaming strategy can effectively reduce the average delivery latency of mobile VR videos when compared with existing caching algorithms.

Key words 360° video; content delivery; mobile edge computing; virtual reality

虚拟现实 (virtual reality, VR) 视频是指全景视频, 如水平 360°×垂直 180°全景视频、水平 180°×垂直 180°全景视频等, 用户可借助 VR 眼镜等虚拟现实设备观看全景视频, 并获得身临其境的视觉感受^[1]。其沉浸式体验对传输速率、时延和可靠性提出很高的要求。传输过程中的延迟和抖动会造成眩晕等症

状, 影响用户体验。在无线网络上传输的 VR 视频也被称为移动 VR 视频。移动 VR 视频作为 5G 无线网络的重要应用, 其传输问题在学术界和工业界引起了广泛的关注。

在进一步分析移动 VR 视频传输问题之前, 先简要了解一种典型的 360°VR 视频的制作、传输与

收稿日期: 2022-06-21; 修回日期: 2022-12-12

基金项目: 四川省科技厅项目 (2021YFHO171)

作者简介: 郑成渝 (1968-), 男, 高级工程师, 主要从事电信网络方面的研究。

*通信作者: 唐勇, E-mail: worldgulit@uestc.edu.cn

播放流程^[1-2]。

1) 制作与编码 VR 视频。一般使用多个摄像机拍摄画面并将抓捕的画面投影到球面空间进行缝合得到全景视频。接着利用等距柱状投影 (equi-rectangular projection, ERP) 将二维球面视频图像映射成二维平面矩形视频图像, 并切割成多个图像块, 然后送进视频编码器进行编码。最终一个完整的 VR 视频文件被转化为多个可独立编码和解码的分块视频文件。

2) 传输 VR 视频。在非直播场景下, 编码后的 VR 视频被缓存在数据中心的内容服务器中。移动 VR 设备在用户观看本地已缓存的 VR 视频过程中, 会提前向内容服务器请求本地未缓存的 VR 视频。被请求的 VR 视频经过互联网络、5G 核心网络、回传网络和前传网络传输最终抵达移动 VR 设备^[3]。

3) 播放 VR 视频。请求的 VR 视频由多个分块视频文件组成。移动 VR 设备对这些分块视频文件解码, 然后将视频图像投影到球面空间, 最后渲染展示到设备屏幕上。

用户使用 VR 设备观看 VR 视频时, 有一定的视角限制, 只能看见 360° 全景视野的一部分画面, 这部分可见区域也被称为视场角 (field of view, FoV)。在流行的移动 VR 设备中, 对于 360° × 180° 的 VR 视频而言, FoV 的大小在 90° × 90° 左右^[4], 仅占全景视野的 12.5%。于是现有传输方案^[5-7]通常仅传输 FoV 视频文件而非整个全景视频, 以降低带宽消耗和请求超时的可能性。FoV 视频文件可以看成由多个分块视频文件组成。为了避免用户突然转动头部出现黑块, 传输 FoV 视频文件通常也指以高码率传输 FoV 内的画面而 FoV 外的画面采用低码率传输或者直接使用纯色填充。

由于核心网络和回程网络充满了不确定性, 移动 VR 设备直接向数据中心的内容服务器请求 FoV 视频文件同样会使得分发时延过大。5G 网络架构中提出的移动边缘计算 (mobile edge computing, MEC) 技术可应用在 VR 传输场景下缓解分发时延过大问题^[8]。一种实践方案是将 MEC 服务器部署在基站附近, 为在基站信号覆盖范围内的移动 VR 设备提供缓存服务。在 VR 视频传输场景下, 将热门 VR 视频缓存到 MEC 服务器中。当移动 VR 设备发出对这些 VR 视频的请求时, 靠近该设备的 MEC 服务器便可以直接处理该请求, 降低请求的视频内容经过互联网和核心网再达到基站的时

延。除此之外, MEC 服务器还可为移动 VR 设备提供计算服务。播放 VR 视频前, 移动 VR 设备需要对视频进行解码和投影操作。虽然这部分计算操作的复杂度比较低, 但并不是所有移动 VR 设备都具备符合需求的计算能力^[9]。对于计算能力不足的移动 VR 设备, 考虑将其计算任务卸载到邻近的 MEC 服务器上处理。

根据思科的最新报告显示, 到 2023 年, 连接到 IP 网络的设备数量将是全球人口的 3 倍以上, 移动数据流量将占到总 IP 流量的 75% 以上^[10]。单 MEC 服务器的缓存能力与数据中心海量存储能力相比十分局限, 而这日益增多的数据量使得这差异愈发明显。此时单 MEC 服务器的优势已不再显著。为满足更多用户的需求, 多 MEC 服务器协作传输 VR 视频正成为一种可行方案。

近年来, 越来越多研究人员也对移动 VR 视频传输问题展开研究。文献 [11] 提出了基于分块来传输 VR 视频的方案, 但没有考虑到利用 MEC 技术来加快传输移动 VR 视频的可能性。文献 [2, 12] 提出了基于 MEC 网络架构来传输移动 VR 视频的方案。文献 [2] 提出了一种基于 FoV 来传输 VR 视频的专用网络架构方案。文献 [12] 把联合优化计算和缓存作为 VR 视频传输的关键问题, 并通过仿真实验证明了方案收益。然而这两项研究都仅在单 MEC 服务器下工作, 没有考虑多 MEC 服务器协作的潜在可能。

本文提出一种基于多 MEC 服务器协作的 VR 视频传输网络架构, 阐述多 MEC 协作缓存和传输 VR 视频的方式。然后以降低 VR 视频平均分发时延为优化目标, 建立混合整数线性规划 (mixed-integer linear programming, MILP) 模型, 通过求解模型得到缓存方案。最后通过实验证明多 MEC 协作缓存和传输 VR 视频可有效降低 VR 视频的分发时延。

1 系统模型

1.1 多 MEC 协作网络架构

基于多 MEC 协作的 VR 视频缓存和传输网络架构如图 1 所示。该传输网络架构由远程内容服务器、5G 核心网络、MEC 协作域 3 个部分组成。远程内容服务器主要负责 VR 视频的前期制作与分发, 包括视频拼接缝合、VR 视频映射和编码, 以及将编码后的视频分发到网络边缘的 MEC 服务器上。

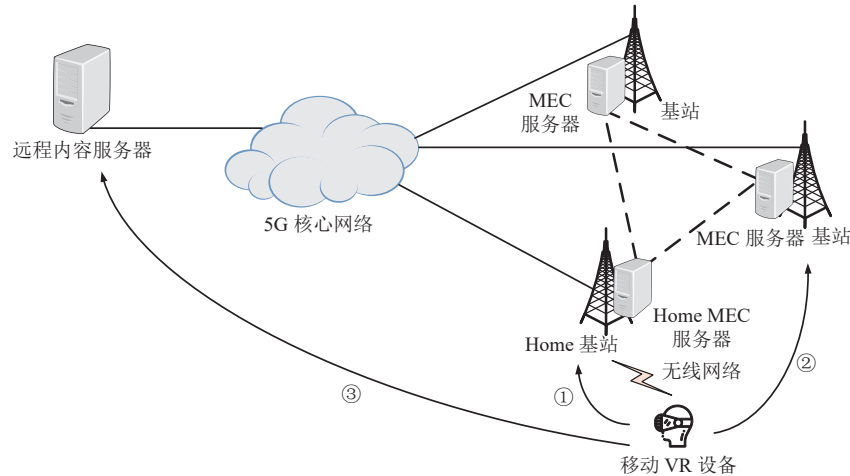


图 1 多 MEC 协作缓存和传输移动 VR 视频架构

在未引入 MEC 的网络中, 被请求的 VR 视频数据在跨越多个网络后才会抵达移动 VR 设备。此传输模式将使得分发时延过大和 5G 核心网络承受较大流量压力。为弥补此传输模式的不足, 本文在传统网络中引入 MEC 服务器并将其部署在基站 (base station, BS) 附近。单个 MEC 服务器的缓存能力相较于数据中心是十分有限的, 为了满足更多请求, 本文提出多 MEC 协作方式。

因为一个 5G 基站的覆盖范围比较小 ($\leq 500 \text{ m}^{[13]}$), 一个生活/商业区域通常会部署多个基站, 于是可以人为地将这些地理位置相近的多个基站以及部署在其附近的 MEC 服务器划分成一个协作区域, 让其共享存储空间与计算资源。下面简称协作区域为协作域。

多 MEC 协作系统周期性地运行缓存算法, 每次选择协作域中负载最低的 MEC 服务器来计算缓存方案, 并将方案下发给其他 MEC 服务器。收到缓存方案的 MEC 服务器把与方案中一致的已存内容保留, 然后再向其他 MEC 或者远程内容服务器请求方案中的其余内容。每台 MEC 服务器中都维护一张资源访问表, 该表记录了协作域内每台 MEC 服务器缓存的内容。当本地缓存未命中时, MEC 服务器根据资源访问表将请求转发给缓存了目标内容的 MEC 服务器, 让其来处理请求。

本文将与移动 VR 设备通信的基站, 称为该移动 VR 设备的 Home BS, 称部署在 Home BS 附近的 MEC 服务器为其 Home MEC 服务器。在基于多 MEC 协作缓存和传输的网络架构中, 移动 VR 设备请求的视频将由以下 3 条路径之一来响应。路径①: Home MEC 服务器上缓存命中, 由

Home MEC 服务器来直接响应请求。路径②: Home MEC 服务器上缓存未命中, 由 Home MEC 服务器向协作域内缓存了目标视频的 MEC 服务器请求。路径③: 协作域内所有的 MEC 服务器都没有缓存目标视频, 那么 Home MEC 服务器就向远程内容服务器处请求。通过后两种方式获得的视频都将被回传给 Home MEC 服务器, 最终再由它传输给移动 VR 设备。3 条路径对应的内容分发时延是不同的。对于同一请求来说, 路径①的分发时延最低, 路径②次之, 路径③最长。于是通过多 MEC 服务器协作, 尽可能地让请求在协作域内被处理, 会大大降低 VR 视频平均分发时延, 从而提升用户的体验质量。

1.2 缓存和请求模型

多 MEC 协作系统可提供服务的移动 360° VR 视频集合为 $N = \{1, 2, \dots, n\}$ 。这些 360° VR 视频均采用 ERP 方式投影到二维平面。为了便于移动 VR 设备请求和服务器端传输, 二维平面视频被划分成时长固定为 Δt 秒的连续视频片段。每个视频片段按行和列划分成 $i \times j$ 个可独立编码和解码的分块视频文件。每个 VR 视频都支持以两种相差较大的码率来编码。对于 VR 视频 n 来说, 其持续时长为 d_n , 支持低码率 lbr_n 和高码率 hbr_n 。于是得到 VR 视频 n 的大小为 $S_n = d_n \times \text{lbr}_n + d_n \times \text{hbr}_n$ 。

协作域内的 MEC 服务器集合为 $M = \{1, 2, \dots, m\}$ 。MEC 服务器 m 具备 C_m 的缓存空间。一个 VR 视频的所有分块视频文件将存放在同一个 MEC 服务器上。在 MEC 服务器缓存空间限制下得到如下约束不等式:

$$\sum x_{m,n} S_n \leq C_m \quad \forall m \in M \quad (1)$$

式中, 变量 $x_{m,n} \in \{0,1\}$ 表示 VR 视频 n 是否被缓存在 MEC 服务器 m 中。

在用户观看 $[t, t+\Delta t]$ 时段的 VR 视频过程中, 移动 VR 设备预测用户 $[t+\Delta t, t+2\Delta t]$ 的头部运动轨迹并得到预测 FoV, 接着向 Home MEC 服务器请求下一视频片段的 FoV 视频文件。

同文献 [5-7] 的传输方案^[5-7]一样, 本文对构成 FoV 的多个分块视频文件采用高码率传输, 而对构成非 FoV 的多个分块视频文件采用低码率传输。用户在球面空间中的 FoV 投射到二维平面可以近似看成由 $k \times k$ 个分块组成^[14]。于是高码率分块占全景画面的比例为 $w = k \times k / i \times j$, 低码率分块占全景画面的比例为 $(1-w)$ 。服务器端最终需要传输的数据量为 $S_n^{\text{trans}} = w \times \text{hbr}_n \times \Delta t + (1-w) \times \text{lbr}_n \times \Delta t$ 。

1.3 时延模型

传输架构中涉及的通信过程可用概括为 3 段通信过程: 1) MEC 服务器与移动 VR 设备之间的通信, 简记为 MEC-MVRD 通信; 2) MEC 服务器与 MEC 服务器之间的通信, 简称为 MEC-MEC 通信; 3) MEC 服务器与远程内容服务器之间的通信, 简称为 MEC-DC 通信。

在 MEC-MVRD 通信过程中, MEC 服务器所在基站到移动 VR 设备是单跳无线网络, 无线电信号在空气中的传播速度近似于光速, 且移动 VR 设备与基站之间的距离不超过 1 km, 因而它们之间的传播时延可忽略不计。同理, MEC 服务器所在的基站之间通过 X2 接口进行通信^[15], 底层物理链路为光纤, 是单跳有线网络且距离不超过 10 km, 由此 MEC 服务器之间的传播时延也可不用考虑。当 MEC 服务器负载过高时, 数据会排队等待发送。于是 MEC-MVRD 和 MEC-MEC 段通信过程引起的时延主要为排队时延和发送视频内容的时延。

由于云端内容服务器到 MEC 服务器的网络状况比较复杂, 包括多跳有线和无线链路, 精确计算云端内容服务器到 MEC 服务器的通信时延比较困难, 本文使用平均通信时延 $T_{\text{dc} \rightarrow m}$ 来表示 MEC-DC 通信过程的引起的时延。

在本文中, 移动 VR 设备按照是否有投影计算能力分为计算能力不足的 U_1 类移动 VR 设备和计算能力充足的 U_2 类移动 VR 设备。在播放视频之前, U_1 类移动 VR 设备会将投影计算任务卸载到其 Home MEC 服务器上, 从而引起额外计算时延。在 MEC 服务器 m 上执行关于 VR 视频 n 的投影计算

任务而产生的计算时延为:

$$T_{m,n}^{\text{off}} = \frac{S_n^{\text{trans}} \times z_n}{f_m} \quad \forall n \in N, m \in M \quad (2)$$

式中, z_n 表示计算 VR 视频 n 每比特数据所会消耗的 CPU 周期数; f_m 表示 MEC 服务器 m 的 CPU 计算频率。

在实际中, 可配置物理层和更高层来设定所有基站对移动 VR 设备都具备相同的数据传输速率 R_1 , 各基站之间通信具有相同的数据传输速率 R_2 。

从上述分析中可以得到, 从服务器端下发 FoV 视频文件到移动 VR 设备端的分发时延主要由计算时延、排队时延和发送时延构成。由此可总结出两类移动 VR 设备请求关于 VR 视频 n 的 FoV 视频文件经由 3 条路径产生的分发时延。

U_1 类移动 VR 设备的请求通过路径①得到响应引发的分发时延可表示为:

$$T_{U_1}^1 = T_{\text{que}} + T_{m,n}^{\text{off}} + \frac{D(S_n^{\text{trans}})}{R_1} \quad (3)$$

式中, T_{que} 表示请求的内容在协作域内传输过程中等待发送经历的排队时延; $D(S_n^{\text{trans}}) = \alpha \times S_n^{\text{trans}}$ 表示经过计算膨胀后的数据量, 通常 $\alpha \geq 2$ ^[9]。

U_2 类移动 VR 设备的请求通过路径①得到响应引起的分发时延可表示为:

$$T_{U_2}^1 = T_{\text{que}} + \frac{S_n^{\text{trans}}}{R_1} \quad (4)$$

U_1 类移动 VR 设备的请求通过路径②得到响应引起的分发时延可表示为:

$$T_{U_1}^2 = T_{\text{que}} + T_{m,n}^{\text{off}} + \frac{D(S_n^{\text{trans}})}{R_1} + \frac{S_n^{\text{trans}}}{R_2} \quad (5)$$

相比从路径①得到视频内容, 从路径②获取视频内容会多出一个发送时延 S_n^{trans}/R_2 , 即视频内容从协作域内其他 MEC 服务器中被发送出去的时延。同理, U_2 类移动 VR 设备的请求通过路径②得到响应的分发时延可表示为:

$$T_{U_2}^2 = T_{\text{que}} + \frac{S_n^{\text{trans}}}{R_1} + \frac{S_n^{\text{trans}}}{R_2} \quad (6)$$

U_1 类移动 VR 设备的请求通过路径③得到响应引起的分发时延可表示为:

$$T_{U_1}^3 = T_{\text{que}} + T_{m,n}^{\text{off}} + \frac{D(S_n^{\text{trans}})}{R_1} + T_{\text{dc} \rightarrow m} \quad (7)$$

U_2 类移动 VR 设备的请求通过路径③得到响应引起的分发时延可表示为:

$$T_{U_2}^3 = T_{\text{que}} + \frac{S_n^{\text{trans}}}{R_1} + T_{\text{dc} \rightarrow m} \quad (8)$$

从 1.1 节可以得知, MEC 服务器 m 服务范围内的移动 VR 设备向其请求 VR 视频 n 只会从 3 条路径之一得到响应。下面引入 3 个请求路径变量来分别表示 VR 视频 n 的 3 种获取位置。 $y_{m,n} \in \{0,1\}$ 表示 VR 视频 n 是否从 Home MEC 服务器 m 处获取, $y_{m,n}^k \in \{0,1\}$ 表示 VR 视频 n 是否从协作域内 MEC 服务器 k 处获取, $y_{m,n}^{\text{dc}} \in \{0,1\}$ 表示 VR 视频 n 是否从远程内容服务器中获取。因为 VR 视频 n 只能从一个位置获取, 于是存在约束:

$$\begin{aligned} y_{m,n} + y_{m,n}^k + y_{m,n}^{\text{dc}} &= 1 \\ \forall n \in N, m \in M, k \in M \end{aligned} \quad (9)$$

引入请求路径变量后, U_1 类移动 VR 设备向 MEC 服务器 m 请求 VR 视频 n 引起的分发时延可整合表示为:

$$T_{m,n}^{U_1} = y_{m,n} T_{U_1}^1 + y_{m,n}^k T_{U_1}^2 + y_{m,n}^{\text{dc}} T_{U_1}^3 \quad (10)$$

U_2 类移动 VR 设备向 MEC 服务器 m 请求 VR 视频 n 引起的分发时延也可整合表示为:

$$T_{m,n}^{U_2} = y_{m,n} T_{U_2}^1 + y_{m,n}^k T_{U_2}^2 + y_{m,n}^{\text{dc}} T_{U_2}^3 \quad (11)$$

1.4 问题建模

用户获取视频的平均分发时延是一项衡量用户体验质量的重要性能指标^[6]。平均内容交付时延越小, 意味着越多的用户请求能由协作域内的 MEC 服务器满足, 用户的体验质量也就越高。于是本文以最小化协作域内 VR 视频平均内容分发时延 (average delivery latency, ADL) 为优化目标建立 MILP 模型, 简称为 MADL。MADL 具体表示如下:

$$\min L$$

$$\text{s.t. 式(1), 式(9)}$$

$$y_{m,n} \leq x_{m,n} \quad \forall n \in N, m \in M \quad (12)$$

$$y_{m,n}^k \leq x_{k,n} \quad \forall n \in N, m \in M, k \in M \quad (13)$$

$$L = \frac{\sum_m \sum_n \lambda_m^n (r_m^{U_1} T_{m,n}^{U_1} + r_m^{U_2} T_{m,n}^{U_2})}{\sum_m (r_m^{U_1} + r_m^{U_2})} \quad (14)$$

约束式 (1) 确保了每台 MEC 服务器缓存的 VR 视频大小总和不会超过其最大存储空间限制。约束式 (9) 确保了对 VR 视频 n 的请求只能在一个地方得到处理。约束式 (12) 和式 (13) 共同限制了只有

当 MEC 服务器上缓存了 VR 视频内容 n 时, 才能响应对 VR 视频内容 n 的请求。式 (14) 表示模型要优化的目标函数, 其中, $r_m^{U_1}$ 表示 U_1 类移动 VR 设备在 MEC 服务器 m 上的请求到达率; $r_m^{U_2}$ 表示 U_2 类移动 VR 设备在 MEC 服务器 m 上的请求到达率; λ_m^n 表示 VR 视频 n 在 MEC 服务器 m 上的流行程度。

2 实验与结果分析

2.1 实验参数设置

在实验参数设定方面, 本文依据现实物理情况并参考相似研究来设置实验参数值。设定一个协作域内的 MEC 服务器数量为 10, 此协作域的服务范围可完整覆盖一个大型商场^[13], 在实际情况中可根据真实需求适当地增加或者减少 MEC 服务器的数量。每个 MEC 服务器都具备 10 GB 的存储空间, 计算频率为 20 GHz。基站到移动 VR 设备的数据发送速率为 100 Mbps, 基站之间的数据发送速率为 500 Mbps^[17]。在 MEC 服务器上执行计算任务, 每比特 VR 视频数据需要消耗 10 个 CPU 周期^[9], 计算后数据量的膨胀系数 $\alpha = 2$ ^[9]。

系统提供了 500 个不同的 360° VR 视频, 时长为 [1, 5] min, 低码率为 [3, 8] Mbps, 高码率为 [15, 32] Mbps。以 $\Delta t = 1$ s 将 360° VR 视频经过 ERP 映射后的平面视频分段, 每个视频片段被切分成 4×8 个分块, FoV 平均由 2×2 个分块组成^[14]。假设 360° VR 视频的流行度服从偏斜系数 $\gamma = 0.56$ 的 Zipf 分布^[16], 偏斜系数 γ 表示内容流行度分布的偏斜程度。 U_1 和 U_2 类移动 VR 设备在每台 MEC 服务器上的请求分别服从到达率为 50 和 100 的泊松分布^[3]。通信过程中的排队时延为 [1, 3] ms, 远程内容服务器传输移动 VR 设备请求的 FoV 视频文件到协作域内 MEC 服务器上的通信时延为 [50, 100] ms^[18]。

2.2 性能对比方案和指标

将本文的 MILP 模型与另外两种传统缓存算法 (Distributed 算法和 Self-Top 算法^[3]) 进行性能对比。在 Distributed 算法中, 协作域内每台 MEC 服务器均缓存完全不同的 VR 视频, 以实现在协作域内缓存尽可能多的内容。在 Self-Top 算法中, 协作域内每个 MEC 服务器都缓存其服务范围内最热门的 VR 视频。由于地域相似性, 一个协作域内的用户对 VR 视频的偏好是比较近似的, 于是在 Self-Top 算法下每台 MEC 服务器上缓存的内容具有很高的相似度。

对于 MILP 模型, 本文实验使用商业软件

AMPL/Gurobi 进行求解, 版本号为 9.5.0, 模型求解精度为 $MIPGAP=0.005$ 。

对于性能分析, 主要以协作域的平均内容分发时延 (ADL) 作为算法性能的评估指标。在仿真实验中, 研究系统参数对缓存性能的影响, 包括 MEC 服务器缓存空间大小、MEC 服务器计算频率、VR 视频数量、Zipf 分布的偏斜系数 γ 和移动 VR 设备请求达到率。

2.3 结果分析

图 2 描述了在单 MEC 服务器缓存空间变化下 3 种缓存算法在 ADL 方面的性能对比。从图 2 中可以看出, 所有缓存算法的 ADL 都随着 MEC 服务器缓存空间的增大而减小。这是因为随着缓存空间的增大, 更多的 VR 视频可以被缓存在 MEC 协作域中, 从而更多的请求可以在协作域内直接得到响应。实验结果表明, MADL 具有较低的 ADL。当协作域内 MEC 服务器的缓存空间不足以缓存下所有 VR 视频时, MADL 较于 Self-Top 和 Distributed 的性能表现更加优越。例如, 在单 MEC 服务器缓存空间从 10 G 到 40 G 变化时, MADL 下的 ADL 比 Self-Top 和 Distributed 分别低约 17%~55%, 11%~16%。

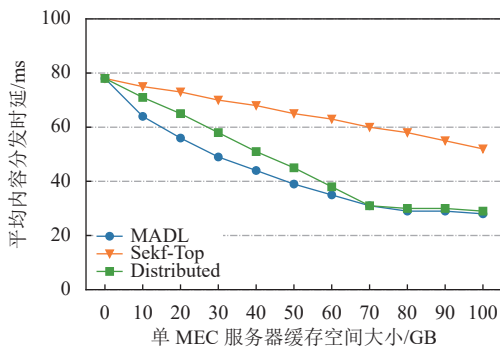


图 2 MEC 服务器缓存大小变化下缓存算法性能对比

图 3 描述了在单 MEC 服务器 CPU 计算频率变化下 3 种缓存算法在 ADL 方面的性能对比。从图 3 中可看出, 所有缓存算法的 ADL 一开始随着 MEC 服务器的 CPU 计算频率增大而减小, 随后趋于稳定。这是因为 U_1 类移动 VR 设备将计算任务卸载到 Home MEC 服务器上产生的计算时延会随着服务器的计算能力增强而降低。当 MEC 服务器的计算能力提升到一定程度时, 计算时延对最终的分发时延不会造成重要影响。由数值结果可以看出, 当 MEC 服务器的 CPU 计算频率从 5 GHz 到 50 GHz

变化时, MADL 下的 ADL 一直低于 Self-Top 和 Distributed, 且差值分别稳定在 16% 和 10%。

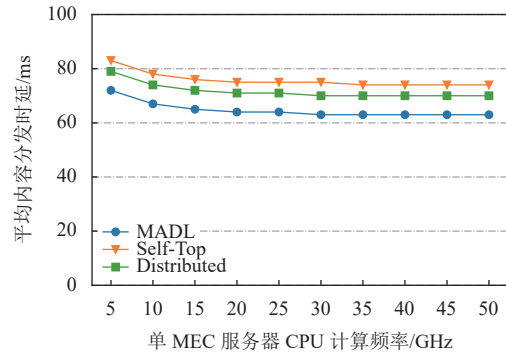


图 3 MEC 服务器 CPU 计算频率变化下缓存算法性能对比

图 4 描述了在 VR 视频内容数量变化下 3 种缓存算法在 ADL 方面的性能对比。从图 4 中可看出, 3 种缓存算法的 ADL 都随着 VR 视频内容数量的增加而增大。因为 MEC 服务器的缓存空间有限, 随着视频内容的不断增多, 越来越多的请求无法从协作域中得到满足。实验结果显示, 当视频内容数量从 100 到 500 变化时, MADL 的 ADL 比 Distributed 低约 7%~11%。在视频内容数量为 100 时, MADL 的 ADL 比 Self-Top 低 67%, 当视频内容数量为 500 时, MADL 的 ADL 比 Self-Top 低 18%。

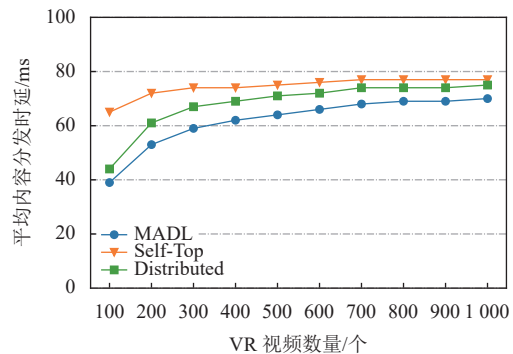


图 4 VR 视频内容数量变化下缓存算法性能对比

图 5 描述了在 Zipf 分布的偏斜系数 γ 变化下 3 种缓存算法在 ADL 方面的性能对比。从图 5 中可看出, 所有算法的 ADL 随着 γ 的增大而减小并且算法之间的性能差距也在逐渐减小。这是因为随着 γ 增大, 流行的视频变得更加集中。此时少量的视频会占据绝大部分的请求流量, 而在协作域中缓存这些少量且热门视频便几乎能满足所有移动 VR 设备的请求。于是出现随着 γ 增大 3 种算法之间的性能差距减小的现象。

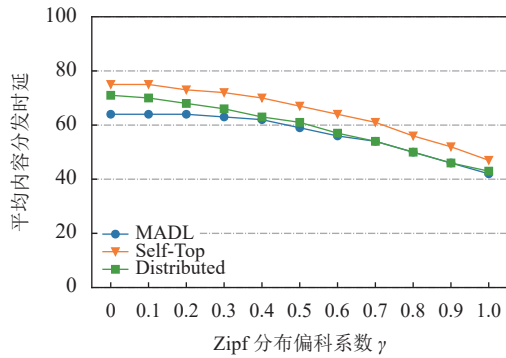
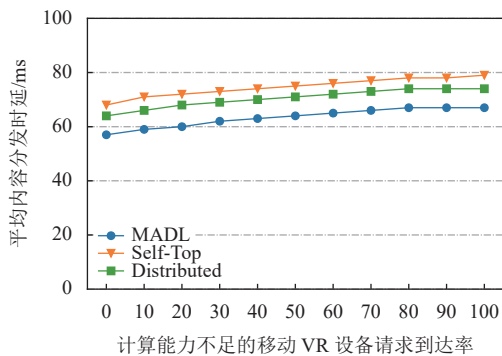
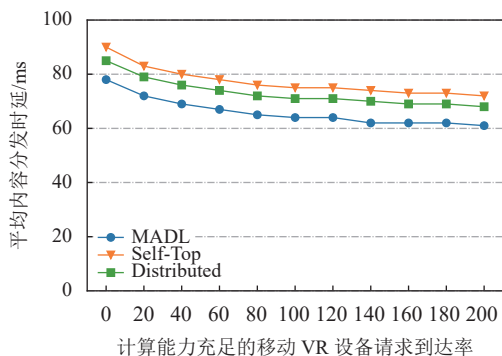
图5 Zipf分布偏斜系数 γ 变化下缓存算法性能对比

图6和图7分别描述了在计算能力不充足的 U_1 类和计算能力充足的 U_2 类移动VR设备的请求到达率变化下3种缓存算法在ADL方面的性能对比。从图6中可以看出,随着 U_1 类移动VR设备请求到达率增高,所有算法的ADL也随之增大。由于 U_1 类移动VR设备会将计算任务卸载到Home MEC服务器上,所以当这类设备的请求数增多时,总分发时延将增大,从而使得ADL增大。从图7中可以看出,随着 U_2 类移动VR设备请求到达率增高,所有算法的ADL却降低。 U_2 类移动VR设备发出的请求不会产生额外的计算时延,所以这类设备的请求数增多,将削弱 U_1 类移动VR设备请求带来的高分发时延影响,因而ADL会降低。

图6 U_1 类VR设备请求到达率变化下缓存算法性能对比图7 U_2 类VR设备请求到达率变化下缓存算法性能对比

3 结束语

为了降低VR视频分发时延以提升用户体验质量,本文提出了一种基于多MEC协作的VR视频缓存和传输网络架构。在考虑协作域内计算能力充足和计算能力不足的两类移动VR设备的请求来决定VR视频的缓存位置的基础上,建立以最小化平均内容分发时延为目标的MILP模型,并通过Gurobi商业求解器得到VR视频的最佳缓存位置。仿真实验证明,本文提出的缓存和传输策略在降低VR视频的平均内容分发时延方面是有效的。

本文研究得到“确定性工业互联网技术基金项目(220481)”支持,在此表示感谢!

参考文献

- [1] 国家广播电视总局. 国家广播电视总局办公厅关于印发5G高新视频系列技术白皮书的通知[EB/OL]. (2020-08-25). [2022-03-17]. https://www.nrta.gov.cn/art/2020/8/25/art_113_52661.html.
- [2] State Administration of Radio and Television. Announcement of the general office of the state administration of radio and television on printing and distributing 5G high-tech video series technical white papers[EB/OL]. (2020-08-25). [2022-03-17]. https://www.nrta.gov.cn/art/2020/8/25/art_113_52661.html.
- [3] MANGIANTE S, KLAS G, NAVON A, et al. VR is on the edge: how to deliver 360° videos in mobile networks[C]// Proceedings of the Workshop on Virtual Reality and Augmented Reality Network. New York, NY: Association for Computing Machinery, 2017: 30-35.
- [4] WANG N, SHEN G, BOSE S K, et al. Zone-Based cooperative content caching and delivery for radio access network with mobile edge computing[J]. *IEEE Access*, 2019, 7: 4031-4044.
- [5] HOU X, DEY S, ZHANG J, et al. Predictive Adaptive streaming to enable mobile 360-degree and vr experiences[J]. *IEEE Transactions on Multimedia*, 2021, 23: 716-731.
- [6] GUNTUR R, OOI W T. On tile assignment for region-of-interest video streaming in a wireless LAN[C]// Proceedings of the 22nd International Workshop on Network and Operating System Support for Digital Audio and Video, New York, NY: Association for Computing Machinery, 2012: 59-64.
- [7] MRAK M, GRGIC M, KUNT M. High-Quality visual experience: Creation, processing and interactivity of high-resolution and high-dimensional video signals[M]. Heidelberg: Springer, 2010.
- [8] KIMATA H, OCHI D, KAMEDA A, et al. Mobile and multi-device interactive panorama video distribution system[C]//The 1st IEEE Global Conference on Consumer Electronics 2012. [S.l.]: IEEE, 2012: 574-578.

- [8] SUKHMANI S, SADEGHI M, EROL-KANTARCI M, et al. Edge caching and computing in 5G for mobile AR/VR and tactile internet[J]. *IEEE MultiMedia*, 2019, 26(1): 21-30.
- [9] SUN Y, CHEN Z, TAO M, et al. Communications, caching, and computing for mobile virtual reality: Modeling and tradeoff[J]. *IEEE Transactions on Communications*, 2019, 67(11): 7573-7586.
- [10] Cisco. Cisco annual internet report-cisco annual internet report (2018-2023) white paper[EB/OL]. (2020-03-09). [2022-03-11]. <https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html>.
- [11] GADDAM V R, RIEGLER M, EG R, et al. Tiling in interactive panoramic video: Approaches and evaluation[J]. *IEEE Transactions on Multimedia*, 2016, 18(9): 1819-1831.
- [12] BASTUG E, BENNIS M, MEDARD M, et al. Toward interconnected virtual reality: Opportunities, challenges, and enablers[J]. *IEEE Communications Magazine*, 2017, 55(6): 110-117.
- [13] VERIZON. How far does 5G reach?[EB/OL]. (2020-04-13). [2022-05-01]. <https://www.verizon.com/about/news/how-far-does-5g-reach>.
- [14] CHENG Q, SHAN H, ZHUANG W, et al. Design and analysis of MEC- and proactive caching-based 360 mobile VR video streaming[J]. *IEEE Transactions on Multimedia*, 2021, 24: 1529-1544.
- [15] NDIKUMANA A, TRAN N H, HO T M, et al. Joint communication, computation, caching, and control in big data multi-access edge computing[J]. *IEEE Transactions on Mobile Computing*, 2020, 19(6): 1359-1374.
- [16] HOU T, FENG G, QIN S, et al. Proactive content caching by exploiting transfer learning for mobile edge computing[J]. *International Journal of Communication Systems*, 2018, 31(11): e3706.
- [17] ZOMAYA. When to use 20 mhz vs 40 mhz vs 80 mhz[EB/OL]. (2019-07-01). [2022-05-02]. <https://www.cbtnuggets.com/blog/certifications/cisco/when-to-use-20mhz-vs-40mhz-vs-80mhz>.
- [18] GSMA. What is the difference in data throughput between LTE-M/NB-IoT and 3G or 4G?[EB/OL]. (2019-10-29). [2022-05-02]. <https://www.gsma.com/iot/resources/what-is-the-difference-in-data-throughput-between-lte-m-nb-iot-and-3g-or-4g/>.

编辑 税红