



研究与开发

# 面向视频流的 MEC 缓存转码联合优化研究

李佳, 谢人超, 贾庆民, 黄韬, 刘韵洁, 孙礼  
(北京邮电大学, 北京 100876)

**摘要:** 为应对未来移动网络所面临的巨大挑战, 业界提出了自适应比特流 (adaptive bit rate, ABR) 技术和移动边缘计算 (mobile edge computing, MEC), 旨在为用户提供高体验质量、低时延、高带宽和多样化的服务。联合 ABR 和 MEC 来优化视频内容分发, 对于提高网络性能和用户体验质量具有重要意义。其中, 各项网络资源的联合优化是重要的研究课题。首先对 MEC 进行了概述, 然后基于面向自适应流的 MEC 缓存转码联合优化问题, 对业界已有工作进行了分析和对比, 并对未来面临的挑战和研究难点进行了归纳和展望。

**关键词:** 自适应视频流; MEC; 缓存; 转码

**中图分类号:** TP393

**文献标识码:** A

**doi:** 10.11959/j.issn.1000-0801.2018222

## A survey on joint optimization of MEC caching and transcoding for video streaming

LI Jia, XIE Renchao, JIA Qingmin, HUANG Tao, LIU Yunjie, SUN Li  
Beijing University of Posts and Telecommunications, Beijing 100876, China

**Abstract:** To deal with the huge challenges in future mobile networks, the industry has proposed adaptive bit rate (ABR) technology and mobile edge computing (MEC), aiming to provide users with diverse services of high quality of experience, low latency and high bandwidth. Combining ABR and MEC to optimize the distribution of video content has been quite important for improving network performance and quality of experience. Especially, the joint optimization of network resources has arisen as an essential research topic. An overview of MEC was firstly given, and then the existing work in the industry on the joint optimization problem of MEC caching and transcoding oriented to adaptive streaming was analyzed and compared. Finally, the existing challenges in the future were summarized.

**Key words:** adaptive video streaming, MEC, caching, transcoding

### 1 引言

随着移动互联网技术的快速发展, 移动网络

数据流量呈现了爆发式的增长趋势, 据 2017 年思科 VNI 报告, 到 2021 年, 全球移动数据流量将达到每月 49 EB (exabyte), 其中移动视频流量将

收稿日期: 2017-12-25; 修回日期: 2018-08-08

基金项目: 中央高校基本科研业务费专项资金资助项目; 国家自然科学基金资助项目 (No.61501042)

**Foundation Items:** The Fundamental Research Funds for the Central Universities, The National Natural Science Foundation of China (No.61501042)

占全球移动数据流量的 78%<sup>[1]</sup>。爆发式增长的数据流量,尤其是视频流量给移动网络造成了巨大的挑战,要求移动网络具有提供更高数据传输速率和更低网络时延的能力,从而为用户提供更好的体验质量 (quality of experience, QoE)。

为了应对上述网络挑战,一方面业界提出了在移动网络中采用自适应比特流 (adaptive bit rate, ABR) 传输技术进行视频分发<sup>[2]</sup>,即在移动网络中将视频文件编码为多种比特率版本,每个版本的视频文件都被切分成多个视频块 (segment),并根据用户设备的能力、网络连接状况和特定的请求,为用户动态选择传送的视频块比特率版本,从而减少视频播放卡顿和重新缓冲率,提升用户的 QoE 体验<sup>[3]</sup>。另一方面,业界也提出了移动边缘计算 (mobile edge computing, MEC),旨在移动网络边缘向内容提供商和应用开发者提供云计算能力和 IT 服务环境,从而为终端用户提供超低时延和高带宽的服务<sup>[4]</sup>。自 ABR 与 MEC 技术提出以后,通过利用 MEC 实现自适应视频流的缓存、转码与自适应分发受到了业界的广泛关注<sup>[5-6]</sup>,并表现出了显著的优势,主要体现在 MEC 在网络边缘提供了存储能力和计算能力,可实现视频内容的就近缓存,并根据动态变化的网络状况在网络边缘进行视频转码,以实现视频的自适应分发,从而降低视频内容的传输时延并节省网络带宽,同时提升用户的 QoE 体验。

目前已有大量工作将 MEC 与 ABR 技术联合考虑,并从不同的部署场景出发,针对视频流的缓存转码等问题进行了研究<sup>[7-18]</sup>,其中不仅包括单 MEC 服务器场景中的资源协同问题,还对分布式多 MEC 之间的资源协作问题进行了讨论,并针对不同的场景和需求,对缓存策略和转码策略进行了设计,从而优化包括系统成本、网络负载和视频容量在内的各项网络指标。另外,由于新型网络架构在增强视频业务质量方面展现的高效性能,部分研究工作还面向视频流分发业务提出了

基于 SDN 的 MEC 架构,从而优化网络架构和视频分发机制。

尽管目前对 MEC 中缓存转码问题的研究已取得了一定的进展,但尚未有系统性的工作对基于 MEC 的自适应视频流分发研究的核心关键技术问题、业界已取得的进展以及未来仍面临的挑战等问题做全面的分析与总结。因此,本文将针对面向视频流的缓存转码联合优化问题,对目前已有工作进行总结和对比,并对目前仍需解决的问题进行归纳和展望。

## 2 自适应视频流场景下的 MEC 部署架构

欧洲电信标准化协会 (European Telecommunications Standards Institute, ETSI) 于 2014 年首次提出了移动边缘计算,并且给出了其“在移动网络边缘提供 IT 服务环境和云计算能力”的定义。MEC 将原本位于云数据中心的服务和功能下沉到网络边缘,提供超低时延、超高带宽的网络环境和实时网络分析能力。由于在回传网和核心网中发挥的低时延保证和容量增强的重要作用,MEC 技术受到了广泛关注,被公认为是 5G 的主要关键技术之一<sup>[19]</sup>。

ETSI MEC 基于虚拟化平台部署,目标是为第三方应用提供标准的体系架构和符合行业标准的 API,使得各种应用可以运行在网络边缘,同时基于 NFV (network function virtualization, 网络功能虚拟化) 的虚拟化基础设施有利于网络运营商的重复利用。MEC 服务器可以部署在网络边缘的不同位置,如 LTE 宏基站 (eNode B)、3G 无线网络控制器 (radio network controller, RNC)、汇聚点或核心网侧等<sup>[20]</sup>。

MEC 服务器通常具有较高的计算能力,适合于分析处理大量数据,可以为 AR 和 VR 等计算密集型业务提供高效的计算资源。另外,由于 MEC 在地理上十分接近用户或信息源,可以显著降低业务响应的时延,并减小回传网和核心网发



生网络拥塞的可能性。最后,位于接入网侧的 MEC 能够实时感知网络数据,包括无线链路状况和用户行为信息及位置信息等,从而进行链路感知自适应,极大地改善用户的服务质量体验<sup>[21-22]</sup>。

与传统的网络架构相比,MEC 具有许多显著的优势,可以有效解决传统网络模式中高时延和低效率等问题。在面向视频流业务时,MEC 的存储计算资源以及网络感知能力可以有效地支持 ABR 技术。一方面,MEC 的分布式边缘存储资源可以对视频内容进行缓存,并且对视频请求进行本地卸载,从而既可以缩短用户到视频内容的距离,降低传输时延,也可以减少视频内容的冗余传输,节省网络带宽并提高能量效率,缓解核心网网络压力,另一方面,在 ABR 中,用户请求的视频比特率可基于网络条件、设备功能和用户偏好自适应调整<sup>[23]</sup>,因此往往需要对视频块进行多个比特率版本的缓存,而一个视频块缓存多种比特率版本会造成较大的缓存成本,因此可以选择在 MEC 中缓存一部分较高比特率视频,对于缓存不命中请求,利用 MEC 的计算能力使用视频转码技术将缓存的视频版本转码为请求的比特率版本,视频转码可以提升缓存资源的利用率,在基于 ABR 的视频分发系统中具有重要的作用,此外,MEC 还可以利用其网络感知能力,对用户和网络信息进行分析,有效支持 ABR 技术。在自适应视频流业务中利用 MEC 缓存流行视频块,并在不同比特率版本之间进行转码,实现在网络边缘对用户请求进行响应,已经被认为是解决 ABR 中视频内容分发的一个重要趋势。

MEC 的缓存和转码能力可以有效支持 ABR 视频业务,从而缓解源服务器压力并提升视频观看质量,但在 MEC 资源容量有限的情况下,缓存转码资源的滥用会造成 MEC 负载过重,降低资源效率,影响用户体验,因此需要对不同视频业务的需求和特点进行分析,在各项资源之间进行权衡和分配,使得视频缓存转码的效率最优。除此

之外,由于 MEC 分布式部署的特点,可以合理利用邻近 MEC 服务器的缓存转码资源,协作处理视频请求,优化整个网络的视频分发质量和资源效率。因此,研究 MEC 中面向视频流的计算、缓存与网络资源的联合优化问题,均衡地对缓存、计算和带宽资源进行合理编排,对优化系统整体性能具有重要意义。

图 1 为一个自适应视频流场景下的 MEC 部署架构,MEC 服务器的位置部署在基站之后,其中部署了缓存和计算资源,可以对视频文件进行缓存和转码,当用户的请求到达 MEC 时,若缓存命中或转码命中,就可以在本地对请求进行响应,如图 1 中的(a)路线和(b)路线,而不需要经过图 1 中(c)路线所示的回传网和核心网的传输,从源服务器获取视频文件。这样一来,既显著降低了用户请求响应的时延,保证用户体验,同时还可以避免网络拥塞,节省回传网和核心网的资源。

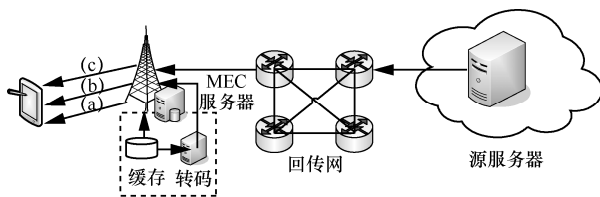


图 1 自适应视频流场景下的 MEC 部署架构

### 3 面向视频流的缓存转码资源联合优化

利用 MEC 实现 ABR 技术,可以有效解决视频内容的缓存转码和分发问题,其中缓存转码资源的联合优化已经被认为是一个重要的研究课题。在传统的视频流分发系统中,缓存和转码都在云端完成,由于传统方案中较高的传输时延和回传压力,MEC 逐渐得到了广泛的应用,研究结果表明部署 MEC 可以有效提升系统性能和用户体验。

#### 3.1 在云端的缓存转码联合优化

媒体云作为当前一种有效实现自适应流分发的框架,在基于云的体系架构中,动态按需编排

虚拟化存储和计算资源,为显著降低系统成本提供了可能性,图2为在媒体云中应用自适应视频流业务的缓存转码示意图,媒体云的流媒体引擎从媒体库中获取视频内容,并进行视频内容的缓存和转码,进而实现自适应视频流向用户的分发。Jin 等人<sup>[7]</sup>研究了媒体云中缓存和转码的最优化策略问题,旨在动态调度各种资源,从而最小化某个节点响应单个请求的总运营成本。具体来说,在缓存策略方面,该方案考虑了视频的流行度,即对流行度最高的若干视频内容缓存全部比特率版本,对流行度次高的一部分视频内容缓存最高比特率版本。当在节点缓存命中时,直接利用缓存响应请求;当转码命中时,比较转码成本和带宽成本,选择成本较小的方式提供服务;当本地缓存不可用时,则从源服务器获取内容。参考文献[7]将缓存、转码和带宽成本之间的权衡问题建模成一个凸优化问题,然后采用两步分析法,分别对缓存资源分配和转码配置策略进行优化,最后得到最优的节点缓存空间大小和所有版本都缓存的视频数量。

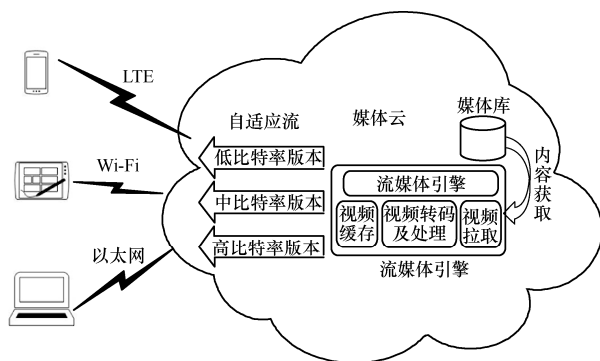


图2 媒体云中面向自适应视频流的缓存与转码

面向视频流的资源联合优化,不仅可以从网络资源配置的角度出发,还可以从用户请求内容的角度进行视频块的调度决策。Gao 等人<sup>[8]</sup>根据对用户视频行为的分析,研究了媒体云中视频内容的管理问题,提出了一种成本高效的部分转码方案。具体来说,每个视频被分为一组片段,其中

一部分片段被预处理并存储在缓存中,而其他片段则在播放过程中被快速转码。缓存策略方面,服务器中存储着源视频文件 and 一部分转码版本,源文件可以转码为任何比特率的版本,另外基于用户实时变化的比特率需求和部分转码的方案,采取动态缓存策略,从而在每个时隙根据用户请求信息来决定是否缓存。当用户请求在节点缓存命中时,用户可以直接获取缓存中的视频内容,进而产生存储成本,缓存不命中时,则进行在线转码,进而产生计算成本。该方案考虑了与存储和计算相关的总成本,建立了一个约束随机优化问题,优化目标为最小化某个服务器的长期总成本,决定某个视频段应缓存还是在线转码,然后利用李雅普诺夫优化框架和拉格朗日松弛法,设计了在线算法进行求解。

通过对用户视频行为的分析可以看出,超过60%的视频只有前20%的内容被用户观看,视频内部存在着不同的流行度,从这一角度出发可以进一步优化网络资源调度。Zhao 等人<sup>[9]</sup>研究了云网络中缓存和转码的均衡问题,基于对视频的分段,考虑了视频段之间的不同流行度以及云中存储和计算价格的大小,对系统中所有视频的放置问题进行了研究,从而决定对每个视频段的哪些版本进行缓存或者转码,将总的系统运营成本最小化。具体来说,首先根据视频内部流行度将视频文件分为多个视频段,通过一个转码权重图描述不同比特率版本之间的转码关系,从而计算出各版本之间的转码成本。基于以上的考虑,提出了一个存储和转码的权衡策略,对流行视频段存储多个或者所有版本,对不流行的视频段存储最高比特率版本,并针对用户请求进行转码。为了避免转码启动时延对视频播放效果的影响,可以在播放视频段时,提前对下一个视频段进行转码。该方案将该权衡问题描述为一个优化问题,优化目标为最小化与存储和转码相关的总成本,并利用启发式分治算法求解,进行视频段某个版本的



缓存和转码决策。

### 3.2 在 MEC 中的缓存转码联合优化

在云端进行缓存和转码,会带来较高的传输时延,增加回传网和核心网的带宽压力,而 MEC 逐步展现出应对视频业务挑战的优势,利用 MEC 的存储和计算能力,可以有效提高网络的整体性能,并显著改善用户体验。其中,MEC 中缓存和转码的联合优化问题已成为一个重要的研究课题。

图3所示为在分布式MEC架构中实现自适应视频流业务的示意,其中,每个MEC服务器都可以实现视频的缓存和转码,分布式部署的MEC服务器之间可以实现协作式的缓存和转码,从而实现资源的高效利用,进一步提升视频分发效率和用户体验。

为解决无线网络中自适应比特流的缓存挑战, Pedersen 等人在参考文献[10-11]中研究了自适应比特流场景中无线缓存和处理的联合优化问题。该方案首先将视频文件分为多个视频块,每个视频块可以按不同的比特率请求。针对 RAN 的缓存挑战,提出在 RAN 部署有限的计算资源,从而可以进行视频块之间的转码,缓解存储压力。基于以上考虑,提出了基于 ABR 感知的主动/被动的联合转码和缓存资源的策略,具体来说,对于视频请求有 3 种内容获取方式,从缓存处直接

获取对应版本,对缓存的高比特率版本进行转码,或者通过回传网络从 CDN 处获取。当用户请求在节点缓存命中时,直接由缓存进行响应,缓存不命中时,可以根据给定可用的缓存容量、处理能力和回传带宽,通过转码资源和回传资源分配算法进行转码和回传决策,当采用回传方式时,采用缓存策略对获取的内容进行缓存。参考文献[10-11]中分别采用了两种缓存方式,分别是 LRU(least recently used) 缓存策略和 P-UPP(proactive user preference profile) 缓存策略。该方案制定了一个优化问题,优化目标为最大化无线网络的视频容量,即服务的并发视频请求数量,并采用启发式算法进行了求解,从而对某个视频块的获取方式进行调度决策。

Wang 等人<sup>[12-13]</sup>提出了一个在线转码和地域分布式交付的联合策略,系统架构中包括多个 CDN 区域,每个区域中包含后端服务器和对等服务器,转码任务在后端服务器中完成。该方案考虑了用户的 CDN 区域偏好、区域的转码版本偏好以及视频请求的用户偏好。首先根据用户的 CDN 区域偏好,即考虑服务器到用户的带宽大小对用户进行重定向,选择提供服务的 CDN 区域,该区域中的对等服务器以循环方式提供服务。另外,根据内容的用户偏好和区域的转码版本偏好来安排转码任务,并选择空闲的 CDN 计算资源进行转

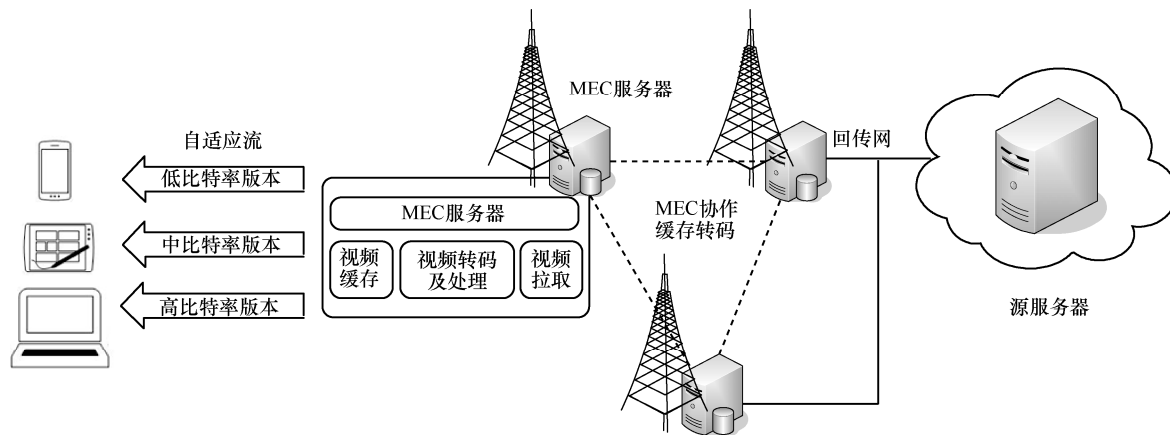


图3 分布式 MEC 架构中面向自适应视频流的协作缓存与转码

码和交付,减少跨区域的复制成本,特别地,根据按需设计的策略,视频段被转码为一组预定义版本,如果转码不及时,可以选择最接近的比特率版本进行转发。最后,该方案对优化问题进行了建模,优化目标为最小化计算成本和复制成本,并采用启发式和分布式算法进行了求解。

在视频业务中,当视频比特率和传输条件不匹配时,会引发网络的拥塞和较高的时延,造成视频播放的卡顿,严重影响用户的观看体验,因此,在资源联合优化的过程中,不仅要从业务的角度出发考虑成本代价,还要从用户的角度出发保障 QoE。众包直播游戏视频流(crowdsourced live game video streaming, CLGVS)是一种新兴的互联网业务,可以使众多异构终端随时随地观看游戏玩家播放的视频,Zheng 等人<sup>[14]</sup>研究了 CLGVS 业务中的在线转码和交付问题,从 CLGVS 服务提供商的角度,通过联合优化动态转码决策、比特率配置和数据中心选择,减少运营成本并保证用户的服务质量。该方案考虑了两种转码策略,分别为门限转码策略和全部转码策略。该问题被建模为一个约束随机优化问题,优化目标包括两部分,即最小化与计算和带宽相关的总运营成本,最大化与时延和比特率相关的用户 QoE。然后利用李雅普诺夫优化框架,设计了一个在线算法 OCTAD (online cloud transcoding and distribution) 进行求解,算法包括 3 个主要部分:动态直播转码决策、自适应比特流配置和智能数据中心选择,从而动态地为每个游戏玩家执行比特率配置,为每个观看者进行转码决策和数据中心选择。同时,为了扩展算法的适用范围,还在设计在线算法时考虑了游戏的类型。

现有的缓存转码优化方案主要基于单服务器独立进行缓存和转码任务决策的场景,没有考虑服务器之间的协作,为了研究在多 MEC 服务器场景下联合优化整体运营成本的问题,Tran 等人<sup>[15-16]</sup>提出了一种移动边缘计算网络中多比特率视频流

的协作缓存和处理策略,称为 CoPro-CoCache。在该方案中,获取视频内容的方法可能有:从本地服务器的缓存中获取,在本地服务器中转码获得,从协作服务器的缓存中获取,在协作服务器中进行转码并传回,从协作服务器传回后在本地进行转码。具体来说,缓存策略方面,本方案不需要内容流行度的先验信息,采用 LRU 缓存策略,将每个小区最流行的视频缓存在对应的基站缓存服务器上,直到缓存存储空间已满,当用户的视频请求需要对缓存中的比特率版本进行转码时,将转码任务分配给负载最小的 MEC 服务器,从而均衡网络负载,这个服务器可以是存储原始版本的 MEC 服务器,即数据提供节点,也可以是交付节点。参考文献[16]将该协作缓存和处理问题建模为一个整数线性规划问题,该问题受存储空间和处理能力的约束,在给定可用资源后,优化目标为对单个视频请求协作制定缓存放置策略和视频调度策略,从而最小化回传网络成本。最后,针对该 NP 问题提出了一个新型的在线算法 JCCP (joint collaborative caching and processing) 来进行求解。

Xu 等人<sup>[17]</sup>提出了 MEC 增强的自适应比特率 (MEC-ABR) 视频传输方案,联合进行缓存和无线资源的分配。在该方案中,MEC 服务器作为控制组件来执行缓存策略并灵活调整视频版本。具体来说,该方案首先考虑了 BS 的流量负载,从而进行 MEC 服务器的存储资源分配,以缓存各 BS 服务范围内的流行视频,并将该存储资源分配问题模拟为 Stackelberg 博弈进行了求解。缓存策略方面,不仅考虑了视频的流行度,还考虑了 RAN 侧的无线信道质量,缓存策略和视频交付可以被灵活地调整,以匹配不同的无线信道。另一方面,该方案将联合缓存和无线资源的分配问题建模为匹配问题,BS 和用户分别根据视频的流行度和无线信道条件维护偏好列表,利用 MEC 的存储和计算能力进行优化,提出了 JCRA (joint cache and



radio resource allocation) 算法来解决这个问题, 并考虑了视频比特率版本的动态调整。

软件定义移动网络 (software-defined mobile network, SDMN)、网内缓存和 MEC 作为下一代移动网络的重要技术, 对增强视频业务质量具有重要意义, Liang 等人<sup>[18]</sup>研究了一个 MEC-SDMN 中的视频速率适应问题, 联合考虑视频速率自适应、带宽配置和 MEC 中的计算资源调度, 设计了一个高效机制。研究目的是在考虑网络资源和视频缓存分布的情况下, 为每个用户找到最佳的视频质量水平。在该方案中, SDN 控制器执行流量管理, 通过最大限度地增大视频的整体平均增益, 提高整个网络的效用, 以帮助用户自适应地选择最佳的视频质量水平。为了最大化 HetNet 的平均视频质量, 该方案制定了一个优化问题, 并采用双分解方法, 将视频数据速率、计算资源和流量管理 (带宽配置和路径选择) 3 部分问题解耦, 独立求解各个变量。

### 3.3 方案对比

本章前两部分主要从云端和 MEC 两个场景出发, 介绍了目前已有工作中面向视频流的缓存、计算和带宽资源的联合优化方案。针对以上方案策略和优化方法等的不同, 下面从缓存策略、转码策略、建模方法和算法等具体方面对已有方案进行了对比, 见表 1。

从表 1 可以看出, 已有工作主要面向的是单服务器架构, 对协作缓存转码问题考虑得较少。缓存方面主要采用一部分全部缓存和一部分缓存最高版本的策略, 在考虑视频流行度的情况下, 可以缓存流行视频的全部版本, 而对不流行视频缓存最高比特率版本。转码策略包括以下几种策略: 缓存不命中直接转码, 与带宽资源做均衡进行转码决策, 协作转码中考虑负载和成本进行转码决策等。另外, 建立优化问题时, 优化目标主要为最小化系统成本, 除此之外, 还包括对视频容量和 QoE 等方面的优化。解决问题的算法主要包括李雅普诺夫优化理论、分析法、博弈论、启

发式算法和在线算法等。

## 4 面向视频流的 MEC 资源优化问题与挑战

在网络边缘部署 MEC 来应对视频流业务, 可以有效降低传输时延并节省网络资源。虽然目前已有大量工作对 MEC 面向 ABR 的缓存、计算和带宽资源的分配和调度问题进行了研究, 但如何综合考虑网络各方面因素, 均衡各项资源, 从而使系统整体性能最优, 有多方面的挑战和研究难点, 本文总结了以下 5 个方面。

### 4.1 缓存转码带宽资源优化与 QoE 优化的均衡问题

QoE 是一种以用户认可程度为标准的服务评价方法, 直接反映了用户在一定客观环境中对适用的服务或业务的整体认可程度<sup>[24]</sup>。自适应流媒体的发展是推动探索增强 QoE 的有效方法的关键驱动力, 从而通过对用户提供差异化服务来保障用户体验<sup>[25]</sup>。MEC 中缓存转码带宽资源的优化与 QoE 优化的均衡问题是一个非常有意义的研究方向, 从视频内容提供商的角度出发, 对于系统的优化一般要考虑两个维度: 一方面要降低缓存、计算和网络的运营成本, 另一方面又要保证终端用户的 QoE。因此, 如何权衡 MEC 缓存转码带宽资源的租赁成本与终端用户的 QoE 保证, 是今后研究的一个重要方向。

### 4.2 缓存转码带宽资源的能量效率优化问题

MEC 的部署将原本位于云端的存储和计算资源下沉到网络边缘, 一方面使得网络边缘可以对用户请求进行响应, 一方面可以减少回传资源的浪费。在 MEC 的网络优化方面, 能量效率问题是关注的重点问题之一。在 MEC 的部署场景中, 内容的缓存、MEC 的计算以及 MEC 之间、MEC 与用户之间的通信都会产生大量的能耗, 从而带来极大的能耗成本。因此, 建立能量高效的资源优化机制, 对缓存、计算和通信资源进行有效的调度, 对于减少系统能耗、提高系统性能有着重

表1 缓存转码联合优化方案对比

参考文献	架构	缓存策略	转码策略	建模	解决问题	算法	偏好考虑
云端	[7]	单服务器	缓存最流行的视频的所有版本, 缓存不流行视频的最高版本	对缓存不命中, 有最高版本缓存, 且转码成本低于带宽成本, 请求进行转码	凸优化问题 优化目标: 最小化与缓存、计算和带宽相关的总运营成本	资源配置: 服务器缓存资源分配和转码策略	两步分析法 视频流行度
	[8]	单服务器	缓存最高版本的源视频文件以及一部分转码版本	部分转码	约束随机优化问题 优化目标: 最小化与存储和计算相关的长期总成本	某视频段的缓存和转码策略	利用李雅普诺夫优化框架和拉格朗日松弛法设计的在线算法 无
	[9]	单服务器	流行视频块缓存多个版本, 不流行视频块缓存最高比特率版本	缓存不命中时进行转码, 还可以提前对下一个视频块进行转码	最小化存储和转码相关的总成本	所有视频段的缓存和转码策略	启发式分治算法 视频段流行度
MEC	[10]	单服务器	LRU 缓存策略	通过转码资源和回传资源分配算法进行转码决策	多背包问题 优化目标: 最大化网络视频容量和 QoE	某视频块的调度策略	启发式算法 无
	[11]		P-UPP 缓存策略				
	[12]	多区域多服务器协作	缓存视频块的最高比特率版本	协作转码	优化目标: 减少计算资源消耗, 最小化复制成本	转码资源的调度	启发式和分布式算法 请求用户偏好、区域用户偏好和转码版本区域偏好
	[13]						
	[14]	多服务器无协作	缓存最高比特率版本的视频	门限转码策略和全部转码策略	约束随机优化问题 优化目标: 最小化与计算和带宽相关的总运营成本, 最大化与时延等相关的用户 QoE	为每个游戏玩家执行比特率配置, 为每个观看者进行转码决策和数据中心选择	利用李雅普诺夫优化框架设计的在线算法 OCTAD
	[15]	多服务器	缓存最流行的视频块, 采用 LRU 缓存策略, 协作缓存	协作转码, 选择负载最小的服务器进行转码	整数线性规划问题 优化目标: 最小化回传网络成本, 受缓存和处理能力约束	某视频块的缓存放置策略和调度策略	JCCP 在线算法 无
	[16]	协作					
	[17]	单服务器	缓存各小区流行视频	按需转码	MEC 缓存资源分配: Stackelberg 博弈 Stackelberg 博弈联合缓存和无线资源分配: 匹配问题	资源分配问题: 缓存资源和无线资源	Stackelberg 博弈论: JCRA 在线算法 视频流行度
	[18]	多服务器	缓存一部分视频的最高版本	根据计算容量进行转码决策	优化目标: 最大化平均视频质量水平	为每个用户寻找最佳视频质量	双分解方法 无

要意义。在面向视频流的 MEC 缓存转码带宽资源的联合优化中, 主要关注缓存能耗、转码能耗和传输能耗, 如何联合考虑 MEC 的计算、转码和传输, 优化提供视频流业务的能量效率是今后研究的一个重点。

#### 4.3 MEC 中基于深度增强学习的缓存和转码

深度增强学习是将深度学习和增强学习结合

起来, 从而实现端对端学习的一种全新的算法, 是通用的人工智能框架, 目前已经成为网络优化的重要方法和工具<sup>[26]</sup>。结合深度增强学习的方法, 对自适应视频流内容进行缓存是一个重要研究方向。在基于 ABR 视频流缓存系统中, 每个视频块都有多个比特率版本, 考虑到边缘网络缓存系统的容量限制, 缓存所有比特率的视频块会造成缓





存资源利用率的降低和网络成本的增加。通过部署在网络无线接入侧的 MEC, 可以实时对网络信息进行感知, 包括网络链路状况和用户行为等数据<sup>[22]</sup>, 利用深度增强学习的方法对这些信息进行分析和学习, 可以预测视频内容的流行度以及用户对响应视频块比特率版本的请求状况, 提前进行资源分配和调度, 对相应视频内容和比特率版本进行缓存, 从而提高缓存命中率和缓存资源的利用率。

#### 4.4 分布式的多 MEC 协作问题

MEC 在边缘网络中的部署, 通常采用分布式的方式, 因此 MEC 带来的缓存和计算资源也分布式的位于网络的不同位置。单个 MEC 的存储空间和计算能力都是有限的, 过多的缓存和计算任务会给 MEC 服务器造成过载, 而回传到云数据中心又会产生较高的回传成本, 因此基于 MEC 分布式的部署方式, 相邻的 MEC 服务器之间可以协作进行缓存和计算, 当前 MEC 服务器没有相应缓存内容或计算资源紧张时, 可以调用其他空闲 MEC 服务器, 此类分布式协作的方式可以有效减少网络运营成本, 提高网络性能<sup>[27]</sup>。因此, 不同 MEC 节点之间如何协作共享资源 (主要包括计算和缓存资源) 成为一个重要的研究问题。例如, 当用户请求的目标视频内容在本地 MEC 服务器没有缓存时, 如何在其他缓存有相应内容的 MEC 节点中选择一个最优的节点; 当本地 MEC 服务器的计算负荷过载时, 如何将本地的计算任务卸载至其他的 MEC 节点, 这都需要 MEC 节点之间的协作。因此研究基于分布式的多 MEC 协作的资源共享机制, 以提高资源的利用率和用户的体验也是今后进行资源联合优化的一个重要方向。

#### 4.5 基于 NFV、SDN 和网络切片等新型技术的资源分配问题

在自适应视频流场景中, 不同的网络环境和用户能力可以动态适配视频流的比特率版本, 而

不同类型的视频业务和不同等级的用户对 QoE 保障有着差异化的需求, 另一方面, 由于单 MEC 服务器的计算和存储资源有限, 分布式多 MEC 场景下的资源协同也面临着很大的需求和挑战<sup>[28]</sup>。如何针对不同业务场景, 利用新型网络技术实现资源的高效管理和分配是 MEC 中面向视频流业务的重要研究课题。网络切片技术可以针对不同应用场景, 将物理网络切割成多个虚拟网络, 从而应对不同场景中对传输时延、移动性、可靠性、安全性以及计费方式的差异性, 利用边缘计算的计算、存储和通信能力, 构建业务所在无线接入网络内的接入网切片, 可以实现业务的本地处理, 缓解核心网压力, 减少传输时延, 改善业务性能<sup>[29]</sup>。除此之外, 未来的 5G 网络还提出了如下演进目标: 基于 SDN/NFV 进行虚拟化, 进行扁平化扩展与增强, 其中 NFV 和 SDN 是实现网络切片的基础, NFV 提供了按需分配的可配置资源共享池, 可以极大地方便资源的统一管理, 同时 SDN 实现了集中式的控制平面, 并通过为用户提供的编程接口, 使用户可以根据上层业务和应用个性化地定制网络资源来满足其特有的需求<sup>[30]</sup>。针对不同的业务场景, 进行有效的网络业务切片和划分, 并通过 SDN 的全局管控和对 NFV 虚拟资源的合理调配, 对优化整体资源效率和网络性能具有重要意义。

## 5 结束语

本文从 MEC 和 ABR 的背景和概述出发, 对目前面向视频流的缓存转码资源联合优化方案进行了介绍和分析, 并主要从缓存策略、转码策略和优化方式等方面对已有方案进行了对比。在对以上方案分析对比的基础上, 研究了面向视频流的 MEC 资源优化问题目前面临的挑战和研究难点, 如与时延优化的均衡问题、能量优化问题和用户行为分析以及分布式 MEC 协作的问题等, 在网络整体优化方面具有重要意义。

## 参考文献:

- [1] Cisco Mobile VNI. Cisco visual networking index: global mobile data traffic forecast update, 2016–2021 white paper[R]. 2017.
- [2] STOCKHAMMER T. Dynamic adaptive streaming over HTTP: standards and design principles[C]//The Second Annual ACM Conference on Multimedia Systems, Feb 23-25, 2011, San Jose, CA, USA. New York: ACM Press, 2011: 133-144.
- [3] YIN X Q, JINDAL A, SEKAR V, et al. A control-theoretic approach for dynamic adaptive video streaming over HTTP[C]//The 2015 ACM Conference on Special Interest Group on Data Communication Pages, August 17-21, 2015, London, UK. New York: ACM Press, 2015: 325-338.
- [4] ETSI. Mobile edge computing—a key technology towards 5G[R]. 2015.
- [5] LI Y, FRANGOUDIS P A, HADJADJ-AOUL Y, et al. A mobile edge computing-based architecture for improved adaptive HTTP video delivery[C]//2016 IEEE Conference on Standards for Communications and Networking (CSCN), Oct 29-31, 2016, Paris, France. Piscataway: IEEE Press, 2016: 1-6.
- [6] WANG C C, LIN Z N, YANG S R, et al. Mobile edge computing-enabled channel-aware video streaming for 4G LTE[C]//Wireless Communications and Mobile Computing Conference, June 26-30, 2017, Valencia, Spain. Piscataway: IEEE Press, 2017: 564-569.
- [7] JIN Y, WEN Y, WESTPHAL C. Optimal transcoding and caching for adaptive streaming in media cloud: an analytical approach[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2015, 25(12): 1914-1925.
- [8] GAO G, ZHANG W, WEN Y, et al. Towards cost-efficient video transcoding in media cloud: insights learned from user viewing patterns[J]. IEEE Transactions on Multimedia, 2015, 17(8): 1286-1296.
- [9] ZHAO H, ZHENG Q, ZHANG W, et al. A segment-based storage and transcoding trade-off strategy for multi-version VoD systems in the cloud[J]. IEEE Transactions on Multimedia, 2017, 19(1): 149-159.
- [10] AHLEHAGH H, DEY S. Adaptive bit rate capable video caching and scheduling[C]//IEEE WCNC'13, April 7-10, 2013, Shanghai, China. Piscataway: IEEE Press, 2013: 1357-1362.
- [11] 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.
- [12] WANG Z, SUN L, WU C, et al. Joint online transcoding and geo-distributed delivery for dynamic adaptive streaming[C]//IEEE INFOCOM'14, April 29-May 1, 2014, Toronto, Canada. Piscataway: IEEE Press, 2014: 91-99.
- [13] WANG Z, SUN L, WU C, et al. A joint online transcoding and delivery approach for dynamic adaptive streaming[J]. IEEE Transactions on Multimedia, 2015, 17(6): 867-879.
- [14] ZHENG Y, WU D, KE Y, et al. Online cloud transcoding and distribution for crowdsourced live game video streaming[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2017, 27(8): 1777-1789.
- [15] TRAN T X, HAJISAMI A, PANDEY P, et al. Collaborative mobile edge computing in 5G networks: new paradigms, scenarios, and challenges[J]. IEEE Communications Magazine, 2017, 55(4): 54-61.
- [16] TRAN T X, PANDEY P, HAJISAMI A, et al. Collaborative multi-bitrate video caching and processing in mobile-edge computing networks[C]//WONS'13, March 18-20, 2017, Banff, Canada. [S.l.:s.n.], 2017: 165-172.
- [17] XU X, LIU J, TAO X. Mobile edge computing enhanced adaptive bitrate video delivery with joint cache and radio resource allocation[J]. IEEE Access, 2017(5): 16406-16415.
- [18] LIANG C, HU S. Dynamic video streaming in caching-enabled wireless mobile networks[J]. arXiv: 1706.09536, 2017.
- [19] 王胡成, 徐晖, 程志密, 等. 5G 网络技术研究现状和发展趋势[J]. 电信科学, 2015, 31(9): 149-155.  
WANG H C, XU H, CHENG Z M, et al. Current research and development trend of 5G network technologies [J]. Telecommunications Science, 2015, 31(9): 149-155.
- [20] 李子姝, 谢人超, 孙礼, 等. 移动边缘计算综述[J]. 电信科学, 2018, 34(1): 87-101.  
LI Z S, XIE R C, SUN L, et al. A survey of mobile edge computing[J]. Telecommunications Science, 2018, 34(1): 87-101.
- [21] TALEB T, SAMDANIS K, MADA B, et al. On multi-access edge computing: a survey of the emerging 5G network edge cloud architecture and orchestration[J]. IEEE Communication Surveys & Tutorials, 2017, 19(3): 1657-1681.
- [22] WANG S, ZHANG X, ZHANG Y, et al. A survey on mobile edge networks: convergence of computing, caching and communications[J]. IEEE Access, 2017, 5(3): 6757-6779.
- [23] TIMMERER C, GRIWODZ C. Dynamic adaptive streaming over HTTP: from content creation to consumption[C]//Proc. of ACM MM'12, Oct 29-Nov 2, 2012, Nara, Japan. New York: ACM Press, 2012: 1533-1534.
- [24] 赵希鹏, 张欣, 杨大成, 等. 基于 QoE 的无线网络资源调度优化研究[J]. 移动通信, 2014(22): 8-13.  
ZHAO X P, ZHANG X, YANG D C, et al. Research on optimization of wireless network resource scheduling base on QoE[J]. Mobile Communications, 2014(22): 8-13.
- [25] LI C, TONI L, ZOU J, XIONG H, et al. QoE-driven mobile edge caching placement for adaptive video streaming[J]. IEEE Transactions on Multimedia, 2017(9): 1.
- [26] HE T Y, ZHAO N, YIN H. Integrated networking, caching and computing for connected vehicles: a deep reinforcement learning approach[J]. IEEE Transactions on Vehicular Technology, 2017, 99(10): 1.
- [27] GHARAIBEH A, KHREISHAH A, JI B, et al. A provably efficient online collaborative caching algorithm for multicell-coordinated systems[J]. IEEE Transactions on Mobile Computing, 2016, 15(8): 1863-1876.
- [28] WANG W, CAO J, ZHANG W. Edge computing: vision and challenges[J]. IEEE Internet of Things Journal, 2016, 3(5): 637-646.
- [29] 项弘禹, 肖扬文, 张贤, 等. 5G 边缘计算和网络切片技术[J].



电信科学, 2017, 33(6): 54-63.

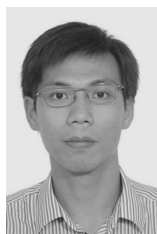
XIANG H Y, XIAO Y W, ZHANG X, et al. Edge computing and network slicing technology in 5G[J]. Telecommunications Science, 2017, 33(6): 54-63.

[30] GPPP E B. QoE-oriented mobile edge service management leveraging SDN and NFV[J]. Mobile Information Systems, 2017(1).

#### [作者简介]



**李佳** (1994-), 女, 北京邮电大学未来网络理论与应用实验室硕士生, 主要研究方向为 5G 网络、移动边缘计算等。



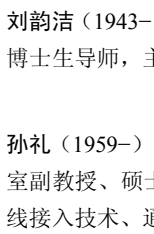
**谢人超** (1984-), 男, 北京邮电大学未来网络理论与应用实验室副教授、硕士生导师, 主要研究方向为信息中心网络、移动网络内容分发技术和移动边缘计算等。



**贾庆民** (1990-), 男, 北京邮电大学未来网络理论与应用实验室博士生, 主要研究方向为新型网络体系架构、内容分发和移动边缘计算等。



**黄韬** (1990-), 男, 北京邮电大学未来网络理论与应用实验室教授、博士生导师, 主要研究方向为新型网络体系架构、内容分发网络软件定义网络等。



**刘韵洁** (1943-), 男, 中国工程院院士, 北京邮电大学教授、博士生导师, 主要研究方向为未来网络体系架构。

**孙礼** (1959-), 男, 北京邮电大学未来网络理论与应用实验室副教授、硕士生导师, 主要研究方向为宽带通信网络、无线接入技术、通信网络交换技术等。