

# 大规模无线网络移动边缘计算和缓存研究

黄永明<sup>1,2</sup>, 郑冲<sup>1</sup>, 张征明<sup>1</sup>, 尤肖虎<sup>1,2</sup>

(1. 东南大学信息科学与工程学院, 江苏 南京 210096; 2. 紫金山实验室, 江苏 南京 211111)

**摘要:** 面向未来 6G 移动通信的大规模网络移动边缘计算与缓存技术, 首先, 介绍了大规模无线网络下移动边缘计算和缓存的架构与原理, 并阐释了移动边缘计算和缓存技术在大规模无线网络中的必要性和普适性。接着, 从计算卸载、边缘缓存、多维资源分配、用户关联和隐私保护这 5 个关键问题出发, 综述和分析了移动边缘计算和缓存赋能大规模无线网络时会引入的新型关键问题以及对应的解决方案研究, 并进一步指出了未来的发展趋势和研究方向。最后, 针对隐私保护问题, 提出了一种基于联邦学习的隐私保护方案, 并通过仿真结果表明所提方案能够同时保护用户数据隐私且改善系统服务质量。

**关键词:** 大规模无线网络; 移动边缘计算; 缓存; 隐私保护; 联邦学习

**中图分类号:** TN92

**文献标识码:** A

**DOI:** 10.11959/j.issn.1000-436x.2021096

## Research on mobile edge computing and caching in massive wireless communication network

HUANG Yongming<sup>1,2</sup>, ZHENG Chong<sup>1</sup>, ZHANG Zhengming<sup>1</sup>, YOU Xiaohu<sup>1,2</sup>

1. School of Information Science and Engineering, Southeast University, Nanjing 210096, China

2. Purple Mountain Laboratory, Nanjing 211111, China

**Abstract:** For the large-scale network mobile edge computing and caching technology of future 6G mobile communications, firstly, the architectures and principles of mobile edge computing and caching in large-scale wireless networks were introduced, and the necessity and universality were clarified. Then, from the perspective of the five key issues in the mobile edge computing and caching enabled large-scale wireless network, including computing offloading, edge caching, multi-dimensional resource allocation, user association and privacy protection, the recent researches and further pointed out the future development trends and research directions were reviewed and analyzed. Finally, for the privacy preservation issue, a federated learning based privacy-preserving scheme was proposed. Simulation results show that the proposed scheme can simultaneously preserve user privacy and improve the quality of service.

**Keywords:** massive wireless network, mobile edge computing, caching, privacy protection, federated learning

### 1 引言

随着 5G 无线通信技术的规模性部署与广泛应用, 通信网络的传输速率、接入密度、频谱效率等都得到了极大的提升, 促进了各类产业与服务的技术变革。但是, 面对未来 Tbit/s 级的数据传输速率、亚毫米级的时间同步、100% 的全球覆盖率、超大规模

接入等更高要求, 5G 将不能满足未来网络需求, 研究人员着眼于 6G 无线通信网络的研究<sup>[1]</sup>。与此同时, 随着移动网络设备及技术的快速升级, 移动接入终端的数量呈爆炸式增长趋势。根据思科公司的白皮书<sup>[2]</sup>统计, 全球移动通信设备数量到 2022 年将会超过 250 亿。因此, 万物互联将是未来网络的必然发展趋势, 也是未来 6G 通信网络的一大重要

收稿日期: 2021-01-21; 修回日期: 2021-03-31

基金项目: 国家自然科学基金资助项目 (No.61720106003); 国家重点研发计划基金资助项目 (No.2018YFB1800801)

**Foundation Items:** The National Natural Science Foundation of China (No.61720106003), The National Key Research and Development Program of China (No.2018YFB1800801)

目标。针对未来万物互联所带来的超密集大规模设备接入技术挑战,面向未来移动网络的大规模组网技术将是一个研究重点。

此外,在超高清视频流分析、智能驾驶,虚拟现实(VR, virtual reality)、增强现实(AR, augmented reality)等新兴业务的驱动下,未来大规模设备在接入网络的同时将会产生大量的计算密集型业务需求<sup>[3]</sup>。这类计算密集型业务通常需要依赖高可靠、低时延的数据实时处理能力来保障服务的正常运行。然而在智能移动设备(SMD, smart mobile device)的计算能力和电池容量受限制的现实情况下,任务的本地执行往往无法满足这些严格的要求。因此,急剧增长的计算密集型业务将会对未来无线网络的实时计算能力提出重大挑战。随着未来大规模组网的逐渐成形,用户无处不在的业务请求与接入将会推动网络中的计算密集型业务需求量达到新的高峰。因此大规模组网在提供大容量、大连接密度技术支持的同时,也将在计算与处理能力方面面临着比传统无线网络更严峻的挑战。

移动边缘计算(MEC, mobile edge computing)技术通过将具有较强计算处理能力的服务器部署在靠近本地设备的小蜂窝基站(SCBS, small cell base station)上来卸载本地SMD的计算压力。作为一个很有前景的接入网范式,MEC被认为是解决未来大规模网络所面临的上述挑战的关键技术之一<sup>[4-8]</sup>。此外,MEC服务器在配置强大算力的同时,也会配备有较强的缓存能力,用来直接满足网络中视频流服务之类的内容服务需求。针对MEC系统中同样强大的缓存能力,移动边缘缓存技术也成为MEC系统中一个重要的研究方向。

MEC服务器在大规模无线网络中的密集部署,可以极大提升网络的计算与缓存能力,但是同样也会引入一系列包括计算卸载问题、多维资源分配问题、用户关联问题、隐私保护问题、边缘缓存问题等在内的亟待解决的新型关键问题。

本文从大规模无线网络中移动边缘计算和缓存技术的基本概念出发,综述了近年来MEC赋能大规模无线网络在计算卸载、多维资源分配、用户关联、边缘缓存、隐私保护这些关键问题上的重点研究工作,并指出了未来的研究方向。同时,针对MEC赋能大规模无线网络中的隐私保护问题,本文提出了一种基于联邦学习(FL, federated learning)的隐私保护方案,实现了在保护用户数据隐私的同

时,提升了系统服务质量。

## 2 背景知识

在未来大规模组网时,通过密集部署的SCBS为用户提供服务,可以缩短终端与业务提供端之间的距离,同时也加强了无线网络在小区边缘的覆盖能力<sup>[9-10]</sup>。SCBS可以通过无线通信链路接入其他SCBS,同样SCBS也可以通过回程链路 with MBS进行连接,各个SCBS之间也可以通过MBS进行间接信息交互。此外,在未来大规模网络接入技术中,设备与设备(D2D, device to device)、车辆与车辆(V2V, vehicle to vehicle)之间的直接通信也将是一种常态。

面向未来移动网络的大规模组网为大连接密度、大流量通信需求提供了保障,但是网络对业务数据的分析和处理能力,以及对内容的提前缓存能力并没有得到明显提升,还是必须依靠远端的云服务器来为用户业务提供计算和存储支持。因此,大规模网络本身在数据处理能力和缓存能力上的缺乏,将在很大程度上限制新型计算密集型、内容服务型业务的应用。鉴于此,移动边缘计算与缓存作为一种很有前景的技术,可被用于打破大规模无线网络计算存储能力不足的瓶颈。

### 2.1 移动边缘计算

欧洲电信标准化协会(ETSI, European Telecommunications Standards Institute)在2014年12月首次提出了MEC技术,该技术的目标是将云计算的功能高效、无缝地集成到无线接入网中,促使在边缘网络端就能够直接实现用户业务的交付<sup>[4]</sup>。之后ETSI于2016年将MEC技术进一步扩展为多接入的边缘计算,从而将边缘计算能力由原本的典型蜂窝网络延伸至包括Wi-Fi接入网、车联网等在内的各类无线接入网络,使MEC技术标准得到进化与完善<sup>[11]</sup>。

在MEC系统中,通过将密集的计算任务下沉到靠近用户端的MEC服务器上,不仅可以极大地缓解核心网的数据传输压力,还能够降低服务的响应时延,提升用户体验。如图1所示,一个移动边缘计算系统的垂直分层架构可以分为终端层、边缘层和中心云层<sup>[12]</sup>。终端层由各类终端通信设备组成,包含智能手机、智能车辆、笔记本电脑、传感器等。终端设备可以借助异构蜂窝通信、车联网通信、物联网通信、大规模组网等多种无线通信技

术接入边缘层，调用边缘网络中的丰富资源。本文接下来将从未来大规模组网接入的角度来探讨未来无线网络下的移动边缘计算与缓存技术。边缘层在各类 MEC 服务器的部署下，拥有了包括通信资源、计算资源、缓存资源和能量资源在内的丰富资源，可以为用户提供强大高效的服务支撑。中心云层位于核心网的云计算中心，通过回程链路与边缘网络连接，以应对边缘层无法满足用户服务的情况。

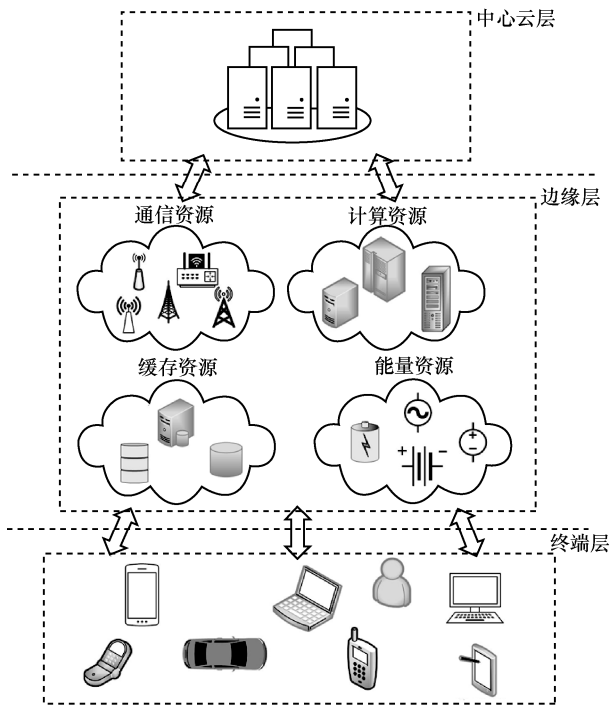


图1 移动边缘计算系统的垂直分层架构

### 2.2 移动边缘缓存

无线终端的普及极大地丰富了用户体验，大量的新型无线业务也应运而生，包括多媒体、Web 浏览、应用程序和社交网络等，这促使传统无线通信转变为大规模无线通信。无线用户对数据服务的渴求进一步加剧了这一现象，这促使无线运营商重新设计当前的网络以适应大规模无线通信，并寻求更先进的技术来提高覆盖率和网络容量，有效地优化用户体验。边缘网络的部署为满足这些前所未有的网络流量需求提供了可能性。边缘网络代表了一种新的基于短程部署、低功耗和低成本的网络范式。利用边缘网络的感知和存储能力，也称为无线边缘缓存<sup>[13-14]</sup>能力，可以实现数据高速传输。将数据内容存储到终端用户附近是实现边缘缓存技术的有效方式。边缘缓存以一种按需应变的方式在网络非

高峰时段提前缓存内容，缓解网络拥挤时的网络拥塞。一方面用户需求是可以预测的，另一方面内存硬件比带宽廉价，而且在无线设备或小型基站上都可以使用。这为边缘缓存技术提供了原理支持和实际部署的可能。边缘缓存系统根据用户请求自适应地调整缓存和多播策略，以满足多个用户的需求，这样既能节约成本又能降低时延。基本的边缘缓存框架如图 2 所示，主要由远端服务器、靠近用户的边缘基站和缓存设备组成。基站从远端服务器预先下载用户可能请求的文件，这些文件被存储在缓存设备中。当用户发出文件请求时，基站根据缓存的文件列表查找请求的文件是否已经缓存在本地。若缓存中已有用户需求的文件，基站直接将该文件多播给提出该文件需求的用户；否则，基站从远端服务器下载未预先缓存的文件，然后再多播给相应的用户。

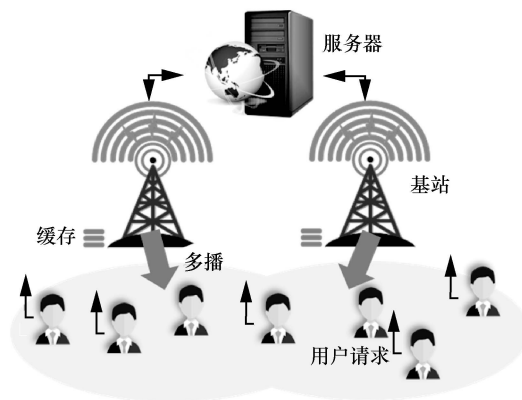


图2 基本的边缘缓存框架

### 3 大规模网络的移动边缘计算与缓存

图 3 是一个典型的 MEC 赋能的大规模移动网络架构。在该系统架构下，网络边缘依靠其大量靠近用户的 MEC 服务器直接在网络边缘上提供各类业务服务来满足业务的高可靠、低时延等需求。然而，虽然 MEC 技术能够有效缓解大规模无线网络中的计算和传输压力，但是也会引入一些新的挑战，需要被进一步研究和解决。

#### 3.1 计算卸载问题

很明显，在 MEC 赋能的大规模无线网络中，MEC 服务器和本地的 SMD 都具有计算的能力，每个计算任务都既能在本地的 SMD 上执行也能上载给 MEC 服务器进行计算，还可以两者协同。区别在于 SMD 受限于自身的体积以及电池容量等硬件

配置，其计算能力往往远小于 MEC 服务器的计算能力。计算卸载则是将本地设备产生的计算任务的全部或部分卸载给其所连接的一个或多个 MEC 服务器上执行，服务器则通过其强大的计算能力来满足用户的计算需求，同时降低计算时延和本地的电池能量损耗。具体地，计算任务在本地设备上执行时，有限的计算能力将会导致计算时延较长，且本地因计算所带来的能耗也会较高，使 SMD 的待机时间受到影响。不过由于计算直接在本地设备上执行，因此传输时延将会很小，网络的传输压力将会减轻。当计算任务卸载到 MEC 服务器上执行时，服务器强大的计算能力将能极大地降低任务的计算时延，同时也能减少本地设备的计算能耗。但是计算任务的上传及计算结果的下发都会带来额外的传输时延和传输能耗，这对网络的传输能力提出了挑战。因此在权衡用户体验、本地设备能耗、网络传输能力和计算能力等多方面因素时，任务的计算卸载策略设计，包括是否卸载、卸载多少等，就直接影响着 MEC 赋能的大规模无线网络的系统性能，是一个亟待解决的挑战。

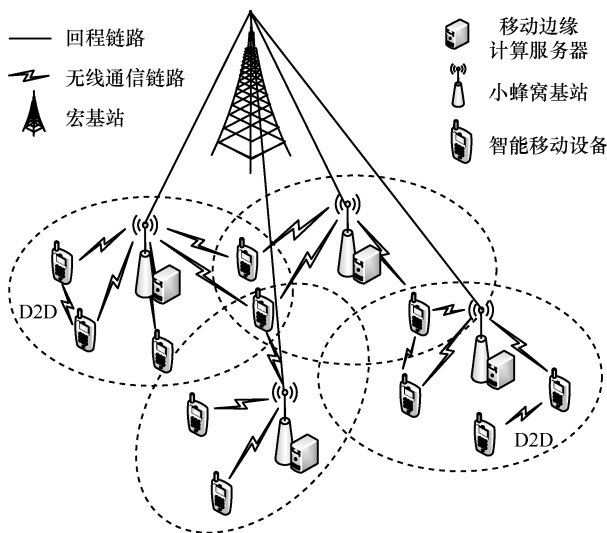


图3 MEC 赋能的大规模移动网络架构

### 1) 能耗敏感型计算卸载

在 MEC 赋能的大规模无线网络中，能耗主要来源于 SMD 的计算能耗，以及任务卸载与回收时的无线传输能耗。其中，降低 SMD 能耗可以有效提升设备的待机时间，而降低网络的传输能耗可以有效提升网络的能量利用效率、降低服务提供商的运营成本。

文献[15]以能效为目标，提出了一种在满足用

户最大容忍时延的前提下最小化 SMD 能耗的卸载策略。在该文献中，用户  $i$  的计算卸载策略被定义为一个二进制的决策变量  $y_i$ ， $y_i=0$  表示用户  $i$  的计算任务将在本地设备上被执行， $y_i=1$  表示计算任务将被卸载到用户  $i$  所连接的 MEC 服务器上被执行。基于此，文献[15]在时延约束的条件下，通过对各用户的卸载策略的联合优化来最小化本地设备的能耗之和，从而得到在单 MEC 节点下基于能效的计算卸载解决方案。然后作者将上述问题抽象为一个非凸混合整数规划问题，并考虑其复杂性，提出了一种基于连续凸逼近 (SCA, successive convex approximation) 的迭代算法来提供原问题的近似解。实验数据表明，该计算卸载方案能在时延约束下有效地降低网络中的设备总能耗。文献[16]针对任务可分割的应用程序，研究了时延约束下能耗最小化的部分卸载方案，并分别提出了一种高复杂度的基于阈值优先级的最优计算卸载算法，以及一种以牺牲部分能耗为代价的低复杂度次优计算卸载算法。仿真结果表明，次优计算卸载算法相对于最优计算卸载算法增加了 20% 左右的能耗来获取更低的算法复杂度。此时，卸载决策变量  $y_i$  不再是一个二进制的 0、1 变量，而是一个 0~1 的连续变量。此时  $y_i$  可视为决策出的卸载给 MEC 服务器进行计算的任务卸载比例系数。

### 2) 时延敏感型计算卸载

除了能耗指标，服务时延同样也是衡量用户体验 (QoE, quality of experience) 的一个至关重要的指标。在计算任务本地执行的情况下，服务时延即为本地设备进行计算所花费的时间。在将计算任务卸载到 MEC 服务器上执行时，服务时延则包含任务上传的传输时延、MEC 服务器进行任务计算的计算时延和计算结果传回用户本地的传输时延。

为了降低时延、保障用户的 QoE，文献[17]在同时以服务时延和任务失败的执行成本作为性能指标的情况下，研究了一种带有能量收集装置的绿色单节点 MEC 系统，并提出了一种有效的计算卸载策略。通过对这 2 个方面指标的联合优化实现了在降低服务时延的同时，提高系统的可靠性，减小任务卸载失败的风险。作者通过分析，将上述联合指标最小化问题建模成一个高维的 Markov 决策问题，并提出了一种基于 Lyapunov 的低复杂度在线动态计算卸载 (LODCO, low-complexity Lyapunov optimization based dynamic computation offloading)

算法来对每个时隙内的计算任务进行卸载决策。同时,在每个时隙内,当卸载决策决定计算任务在本地设备上执行时,所提 LODCO 算法还要联合优化出本地设备中央处理器(CPU, central processing unit)计算频率的分配策略。当卸载决策决定将计算任务卸载到 MEC 服务器上执行时,LODCO 算法则需要联合优化出传输功率的分配策略。文中实验结果表明,LODCO 算法能够有效地降低 64% 的服务时延。

### 3) 时延-能耗敏感型计算卸载

能耗和时延这两大性能指标共同影响着系统的性能和用户的 QoE。因此,也有文献研究了能耗和时延权衡优化的计算卸载方案。文献[18]提出了一种能量感知的计算卸载方案,在能量和时延的限制下,通过最优化 SMD 能耗和服务时延的加权和,实现 SMD 的能量消耗和任务延时之间的权衡。此外,作者还将 SMD 的剩余电量引入能耗-时延权衡的权重因子当中,来提高 SMD 的待机时间,并在此基础上提出了一种结合内部惩罚函数和 D.C. (IPDC) 规划的迭代搜索算法,以获得最优计算卸载决策和资源分配方案。仿真结果表明,所提算法能获得更低的时延能耗权重和,并能明显提高 SMD 的待机时长。文献[19]同样考虑了计算卸载的时延和能耗的权衡问题。作者针对时延敏感型计算任务,在满足容忍时延以及 SMD 电池剩余电量的约束条件下,通过最小化时延能耗加权和的方式来联合优化计算卸载比例、本地设备计算频率以及卸载的数据传输功率。通过对建模出的时延-能耗权衡问题的分析,作者将该高复杂度的非凸且非光滑的权衡问题转化为一个光滑的双凸问题,在此基础上提出了一种基于交替凸搜索(ACS, alternate convex search)的求解算法,大大降低了计算量,并在不同性能参数的数值仿真下展现了先进的性能。

## 3.2 边缘缓存问题

基本的缓存策略有 2 种,一种是在基站端部署的主动缓存,另一种是部署在用户端的编码缓存。为区别于编码缓存,主动缓存也通常被称为非编码缓存。主动缓存是指以主动方式将数据内容提取并存储在基站的缓存空间中。该技术可以将流行度较高的内容放置在离终端用户更近的位置,并在用户发送数据请求的时候直接发送给用户,这可以有效地减少时延。Maddah-Ali 等<sup>[20]</sup>提出了一种编码缓存方案,该方案根据所有用户可用的累积存储容量来

提高缓存和多播增益。

### 1) 非编码缓存

部署在基站端的非编码缓存通常是基于用户请求的历史数据进行的,该类缓存方案首先分析用户请求的历史信息,提取出和文件内容流行度相关的信息,利用这些信息预测未来一段时间较流行的文件,并采用合理的策略有选择性地将它们预先存储到本地。文献[21]提出一种有效的缓存机制,通过将最流行的内容复制到网络边缘,而不是将其存储在中央站点,从而减轻这些海量带宽需求。文献[22]提出了一种基于网络主动性的缓存方案,利用用户的统计信息发现用户的行为模式,以此预测用户需求,并在用户提出相应请求前预存储一定量的数据。这种主动缓存模式进一步提升了存储能力,给无线传输带来增益。文献[23]将缓存看成数据流量卸载的一种方式,将重点放在存储与带宽的权衡方面,根据存储限制条件,在给定的内容流行度分布条件下缓存内容,并提供了缓存未命中概率的闭合表达式。

上述对于无线缓存的研究往往需要建立精确的缓存优化问题,该优化问题往往需要对无线环境进行精确的建模,或者对用户需求进行准确的估计,然后才能使用相应的优化算法求解缓存问题。然而,在实际系统反映实际环境的数学模型往往是难以建立的,尤其在用户需求动态变化的缓存网络中,传统基于数学建模的缓存方案往往具有较强的先验假设,而随着环境的快速变化,这些先验假设可能不再成立。近些年来,在海量数据和高速计算能力的支持下,深度强化学习算法通过对数据进行分析、特征抽取,使算法获得了很强的推断、判断、数据拟合、决策、识别、优化等能力。当前基于强化学习的缓存方案得到了广泛的关注。

现有的基于强化学习的缓存方案的重点是使基站能够学习未知的系统状态转移函数,并相应地缓存最合适的文件内容。文献[24]考虑全双工密集网络的缓存系统,提出一种新的缓存感知和用户-基站多重关联机制,该机制允许用户与多个基站关联。为了获得最佳缓存和关联策略,该文提出一种基于信干噪比的置信界算法,并仿真验证了该算法的可行性和优越性。文献[25]考虑缓存辅助下大规模密集网络中的能耗问题,并提出一种新颖的缓存策略来改善系统能量利用效率。该研究考虑缓存内容的流行度是动态变化并且未知的,所提缓存策略

是基于深度强化学习算法获取到的。文献[26]研究了5G网络中的新型3D视频主动缓存的问题。首先，将多视角3D视频的主动缓存问题建模为联合视角选择和本地内存分配的马尔可夫决策过程。然后，提出一种基于深度确定性策略梯度的缓存方案。考虑到动作空间受系统状态的影响，文献[26]将动态 $k$ 近邻算法嵌入深度强化学习算法中，以实现在可变大小的动作空间中依旧能正常工作的深度强化学习算法。

## 2) 编码缓存

编码缓存利用用户的本地存储空间来提升多播增益。假设基站缓存了 $N$ 个文件 $\mathcal{W} = \{W_1, \dots, W_N\}$ ，每一个文件都具有单位大小。与该基站相连的 $K$ 个用户组成集合 $\mathcal{K} = \{1, \dots, K\}$ ，假设每一个用户拥有大小为 $M$ 的缓存空间。基于此，编码缓存方案主要分为以下2个阶段。

①配置阶段。每一个文件都被独立地等分为 $F$ 个切片， $W_i = \{W_{i,j} : j \in [1, F]\}$ 。每一个切片的大小为 $1/F$ 。这些切片被放置在用户的缓存实体中，并且放置方式与用户的需求无关。该配置阶段在非高峰时段执行。

②分发阶段。每一个用户随机并且独立地发出文件请求，构成向量 $\mathbf{d} = (d_1, \dots, d_K)$ ，其中 $d_k \in [1, N]$ 代表用户 $k \in \mathcal{K}$ 请求文件 $W_{d_k} \in \mathcal{W}$ 。一旦基站接收到 $\mathbf{d}$ ，就向用户广播一个大小最多为 $RF$ （ $R$ 为发送速率）包的编码文件，这样每个用户都可以通过配置阶段接收到的内容和分发阶段接收到的数据来恢复出自己所请求的文件。

图4为编码缓存系统框架。该场景包括一个边缘基站、2个用户和2个文件（ $A$ 和 $B$ ）。在配置阶段，用户尚未发出任何文件需求，网络处于非高峰时段，基站向用户1传输文件切片 $A_1$ 和 $B_1$ ，向用户2传输 $A_2$ 和 $B_2$ 。在分发阶段，首先2个用户分别向基站提出文件请求，假设用户1和用户2的需求分别为文件 $A$ 和文件 $B$ 。当基站收到该请求后，基站将向用户广播 $A_2 \oplus B_1$ ，其中 $\oplus$ 为异或（XOR）操作。用户接收到广播的编码数据后，利用异或运算的性质即可解码出各自需要的文件。

上述编码缓存方案得到了研究人员的广泛关注。文献[27]提出了一种新的分层编码缓存（HCC, hierarchical coded caching）方案。使用HCC，可以在恒定的间隙内实现最佳通信速率。在文献[28]中，

针对通信限制为单跳的无线D2D网络，作者提出了一种基于确定性配置策略的缓存方案，以及一种编码分发策略。在这种编码缓存方案中，用户相互发送线性编码消息以共同满足其需求。结果表明，D2D网络的空间复用增益与单个基站传输的编码多播增益在数量级上相等。随着对编码缓存研究的深入，研究人员注意到为了实现上述编码缓存方案，必须将每个文件分成 $F$ 个切片。切片的数量通常随用户数量呈指数增长，减少 $F$ 可使编码缓存更适合实际系统。为了降低 $F$ ，文献[29]提出了一个配置分发阵列（PDA, placement delivery array）来描述配置和分发这2个阶段。

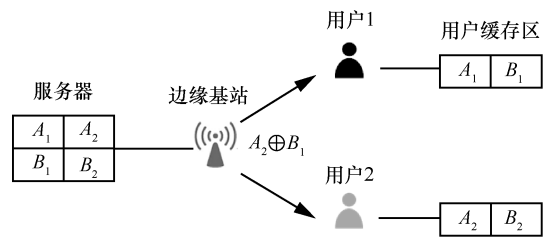


图4 编码缓存系统框架

**定义1** 配置分发阵列<sup>[29]</sup>。对于正整数 $K$ 、 $F$ 和非负整数 $Z$ 、 $S$ ， $F \geq Z$ ，如果满足下列条件，则由符号 $*$ 和 $S$ 个整数 $1, 2, \dots, S$ 组成的大小为 $F \times K$ 的阵列 $\mathbf{P} = (p_{i,j})$ ， $i \in [1, F]$ ， $j \in [1, K]$ 是一个 $(K, F, Z, S)$ 配置分发阵列。

**条件1** 符号 $*$ 在每一列中出现 $Z$ 次。

**条件2** 对于任意2个不同的元素 $p_{i_1, j_1}$ 和 $p_{i_2, j_2}$ ，当且仅当它们排列在不同的行和列中，且 $p_{i_1, j_2} = p_{i_2, j_1} = *$ ，有 $p_{i_1, j_1} = p_{i_2, j_2} = s$ 为一个整数。

根据一个确定的配置分发阵列 $\mathbf{P}$ ，满足 $M/N = Z/F$ 的编码缓存方案可以描述如下。

①配置阶段。每一个文件都被独立地等分为 $F$ 个切片， $W_i = \{W_{i,j} : j \in [1, F]\}$ ，并且用户 $k \in \mathcal{K}$ 接收并存储切片 $C_k = \{W_{i,j} : p_{j,k} = *, \forall i \in [1, N]\}$ 。

②分发阶段。基站接收到请求序列 $\mathbf{d}$ ，在时隙 $s$ 基站广播编码数据 $\bigoplus_{p_{j,k}=s, j \in [1, F], k \in [1, K]} W_{d_k, j}$ 。

上述编码缓存方案可以扩展到多基站或者大规模网络中去。文献[30]考虑了超密集网络场景下的编码缓存问题，假设每个基站都具备有限的缓存空间，可以预先存储部分文件的数据。该文提出了一种双次编码缓存方案，可以有效保证无线编码缓存场景下较高的传输成功概率。进一步地，考虑动

态缓存和多播调度问题,以实现在内容为中心的无线网络中平均时延和能耗的最小化。将上述随机优化问题表述为具有未知转移概率和状态空间巨大的马尔可夫决策过程。然后,提出了一种深度强化学习算法来解决上述问题。所提算法使用变分自动编码器来充分表征环境状态信息,并使用加权双  $Q$  学习方案来减少  $Q$  函数的方差和过估计。数值结果表明,所提方案增加了成功传输的概率,并且缓存和调度策略可以有效地减少时延和能耗。

本节介绍了非编码缓存和编码缓存在无线边缘网络中的应用,指出了传统非编码缓存可以结合深度强化学习,在系统状态转移概率不可知和环境动态变化等场景中,获取一个近似的最优缓存策略来减少时延和能耗。编码缓存可以充分利用用户的缓存能力,进一步提升边缘网络的多播增益。同时,结合深度强化学习,在边缘网络中设计智能的编码缓存策略和多播调度策略也能进一步提高系统的自适应能力。

### 3.3 多维资源分配问题

不同于传统的无线通信网络,在 MEC 赋能的大规模无线网络中,大量的 MEC 服务器、海量的 SMD、密集的蜂窝网络等特征使整个系统中存在更丰富的计算、存储、能量和通信资源。具体地,计算资源是指每个 MEC 服务器以及每本地智能设备都配备有较强大的 CPU 处理器,能够为密集的计算任务提供算力支持。存储资源则代表着系统中各个设备都具有一定的缓存能力,能够缓存一定的数据用于支持应用程序。能量资源是指 SMD 的电池供能以及 MEC 服务器的电源能量,这些能量可以用于任务的计算以及数据的无线传输。通信资源是指 SCBS 与 SMD 之间、SMD 彼此之间的通信链路容量,以及各个 SCBS 彼此之间的回程链路容量。这些丰富的多维度的系统资源是 MEC 赋能大规模无线网络的一大优势,能够用于满足海量业务的各类需求。而合理地利用这些资源,提高资源的利用效率,将会极大地促进网络的性能,提升用户的 QoE。

#### 1) 计算资源分配

前文提到的计算卸载技术,实际上也是一种计算资源的分配。通过卸载策略合理分配 SMD 与 MEC 服务器上的计算资源进行任务计算,以提升计算资源的利用效率,促进系统性能的提升。此外,

由于 MEC 服务器的计算能力相对于 SMD 来说通常都更强大,因此,对 MEC 服务器上的丰富计算资源进行更细致的分配与利用,对于系统的性能也会产生至关重要的影响。文献[31]从时间维度考虑了有连续到来的计算任务时,如何对 MEC 服务器的计算资源进行合理分配,以实现时延约束下的网络平均数据传输量最小化。在每个时隙,作者考虑了当前 MEC 服务器的 CPU 占用状态,然后针对不同的占用状态,分别提出了 5 种相应的连续任务安排策略,以提高计算效率。在此基础上,作者提出了一种通信受限的 MEC 架构来降低通信资源的消耗,并进一步提出了一种基于 Lyapunov 理论的最优任务安排算法来最小化每个计算任务的平均传输数据量。

另外,文献[32]从空间维度对 MEC 服务器的计算资源进行精细分配,以提高计算资源的利用效率。具体地,文献[32]考虑了在 MEC 赋能大规模无线网络场景下,每个用户以及每个 SCBS 的计算需求往往都是高并发的。因此作者针对这种空间维度并发的计算需求,对 MEC 服务器上的计算资源进行空间维度的分配,以同时满足所有用户以及所有 SCBS 的计算需求。接着,作者提出了一种基于势博弈论(PGT, potential game theory)的计算资源分配方案,该方案在降低能耗的同时可以提高计算资源利用效率。所提方案包括基于 PGT 的功率控制方案和基于线下规划的资源分配方案。功率控制方案旨在寻找一组能使 MEC 网络的势函数最大化的基站传输功率,资源分配方案则是根据功率控制方案的结果来使 MEC 网络的平均计算资源分配系数最大,也即使计算资源利用效率最大。数值仿真结果表明,所提方案能显著提升网络中的计算效率和能效。

#### 2) 缓存资源分配

缓存资源作为 MEC 赋能的大规模无线网络中的又一大资源,存在于网络中所有具有缓存能力的设备当中。针对密集的计算任务,提前缓存好所需的计算输入文件,或是缓存下某些被频繁请求的计算任务的计算结果,能够有效地降低服务时延,提高用户的 QoE。针对大规模网络中海量的业务需求,网络中缓存资源的合理利用与分配则具有重要意义。

文献[33]通过研究网络中用户之间的社交关系以及负载的统计数据,提出了 2 种缓存配置方案来合理利用系统的缓存资源。这 2 种方案都能基于客

户之间的社交关系来主动地在客户端上创建访问频率高的内容副本,以提高缓存的命中率,提升缓存资源的使用效率。但是由于文献[33]所提缓存策略是一种静态的缓存配置方案,当网络环境以及用户兴趣发生变化时,这种静态的缓存配置方案将很难继续保持高缓存资源利用效率。文献[34]则考虑到大规模移动网络下的动态场景,提出了 MEC 服务器缓存随着时变的任务请求进行在线更新的动态缓存方案。在文献[34]中,作者为了最小化长期累计服务时延,针对串行前端传输模式和流水线前端传输模式,分别提出了在线缓存交付模式,以及响应式和主动式在线缓存模式。文中通过理论分析,找到了 2 种缓存模式下的最优累计服务时延上界,并通过数值仿真展示了 2 种动态缓存方案的优越性,同时也进一步证实了理论分析的正确性。

### 3) 能量资源分配

在 MEC 赋能的大规模无线网络中提供新型业务服务时,任务的计算与分析、数据的传输、设备之间的通信等都需要消耗能量。对网络中能量资源的合理分配,能够大大提升能效,降低系统能耗。文献[35-36]都对 MEC 网络中的无线传输能耗进行了研究与优化。文献[35]研究了单用户单 MEC 服务器场景下的下行传输功率分配方案设计。首先为了表征基本性能极限,作者假设信道状态信息(CSI, channel state information)和任务状态信息(TSI, task state information)的完备知识(即任务到达时间和数量)是已知的,从而考虑传输功率的离线优化。在这种情况下,作者利用凸优化技术得到了能耗最小化问题的最优解。之后,作者受上述结构化离线能量资源分配方案的启发,考虑了在 CSI/TSI 知识仅为因果已知的情况下,针对能量资源分配和任务卸载的联合优化,开发了启发式在线能耗优化算法。文献[36]则将文献[35]的单用户场景扩展到了多天线多用户的 MEC 系统。在该系统中,作者同样从离线优化和在线优化 2 个方面,对下行传输功率分配方案进行了研究,并分别得到了基于拉格朗日对偶法的离线优化问题的半封闭式最优解和基于滑动窗口以及序列优化的在线能量资源分配方案。仿真结果表明,所提方案相对于一些传统能耗优化方法能够降低 45%~90%的系统能耗。

### 4) 通信资源分配

受无线网络中的通信带宽、通信子载波个数的限制,无线网络中通信资源块分配方案的优劣直接

影响着系统的通信效率以及服务时延。而在 MEC 赋能的大规模无线网络中,高并发的任务请求、连续不断的计算任务卸载、低时延的数据处理需求等都依赖于一个高效的通信资源分配方案来提供保障。因此,对 MEC 赋能大规模无线网络中的通信资源分配方案进行优化是一个提高系统性能的有效手段,同时也是一个极具挑战的难题。

文献[37]针对 MEC 系统中无线虚拟现实业务的通信资源分配问题进行了研究与分析。作者考虑了无线虚拟现实视频业务,充分利用该业务中传输数据的空间相关性来更好地进行上下行传输资源块的管理与分配,进而减少上下行链路的通信负荷。在下行链路中,MEC 服务器可以根据用户的数据相关性,从原始的 VR 视频内容中提取特定的用户感兴趣的视野内容传输给用户,从而减少内容重复传输所带来的传输冗余。在上行链路中,每个 SCBS 可以与具有相似追踪信息的用户关联,从而减少用户视野追踪数据的数据量。具体地,在文献[37]中,该数据相关性感知的传输资源分配问题被建模成一个最小化用户成功传输概率(定义为传输时延小于最大服务容忍时延的概率)的优化问题,并提出了一种基于联合回声状态网络和迁移学习的机器学习求解算法。通过智能地在 SCBS 之间迁移信息,所提算法能快速应对由于用户内容请求和内容请求分发变化而导致的无线网络环境的更改。实验表明,该基于机器学习的智能算法相对于传统算法以及不考虑数据相关性的传输资源分配算法,能够提升 15.8%~29.4%的用户成功传输概率。

### 5) 混合资源分配

通过对上述研究工作的分析可以发现,多维度资源之间往往存在复杂的耦合关系。例如,在进行计算任务卸载时,为了节省小基站与用户之间的通信链路资源而选择在本地设备上进行任务计算,那么本地设备的能耗将会增大。为了节省本地设备的电池电量,选择将计算任务卸载到 MEC 服务器端进行计算,任务的上传以及计算结果的下载都需要消耗传输能量,同时 SCBS 与 SMD 之间通信链路资源的消耗也相应增加。另外,当 SCBS 与 SMD 的缓存资源减少时,通信链路所需的内容传输次数将会增加以保证业务的正常进行,因此通信资源的消耗得到相应增加,同时用于文件传输的次数增多也将导致系统能耗的增加。由此可见,MEC 赋能的大规模无线网络中各类资源之间是相互耦合,彼此

关联的。因此, 如何对这些混合资源进行高效而又快速的联合调配来提升整个无线网络的性能, 是另一个需要被广泛研究的课题。

文献[38-39]分别从机器学习和传统优化的角度, 研究了 MEC 系统中多维资源的联合优化问题。文献[38]将 MEC 系统中的通信与计算资源联合分配问题建模为一个复杂的决策过程, 然后通过数学分析, 设计了一种基于深度强化学习 (DRL, deep reinforcement learning) 的求解算法来解决该多参数优化难题, 以实现最小化资源消耗和计算时延的目标。实验结果表明, 该深度强化学习算法能在通信资源与计算资源的利用效率上实现明显提升。文献[39]则利用设备到设备 (D2D, device-to-device) 通信来实现用户协作以及减少 MEC 服务器负载。作者通过联合优化计算资源、能量资源和通信资源的分配来最小化系统中所有用户的能量消耗和时延的加权总和。对于计算资源分配, 作者提出了一种自适应优化算法来寻找最优解; 对于能量资源分配, 作者提出了一种基于粒子群算法的功率分配算法; 对于通信资源分配, 作者提出了一种基于 Pareto 改进和交换操作的一对一匹配算法, 并将这种一对一匹配算法扩展到多对一匹配场景。最后, 基于上述单独优化方案, 作者从传统优化的角度出发, 提出了一种基于调度的联合计算、功率和通信资源分配算法来实现系统能耗与服务时延的联合优化。

### 3.4 用户关联问题

如图 3 所示, 在大规模无线网络中, 由于大量 SCBS 的密集部署, 使有些 SMD 会同时存在于 2 个甚至多个 SCBS 的覆盖区域内, 这就导致这类用户在接入无线网络时, 拥有多个可供选择的接入节点。在 MEC 赋能的大规模无线网络中, 对于这种用户多 MEC 节点关联问题, 关联准则可以基于不同 SCBS 与用户之间的信道链路状态好坏, 也可以基于不同 SCBS 所配备的 MEC 服务器的计算能力强弱, 还可以基于不同 MEC 服务器中缓存内容的匹配度等。考虑到不同的目标时, 例如时延最小化、能耗最小化和接入数量最大化等, 上述各角度的关联准则会相互制约, 需要进行权衡和综合考虑。此外, 考虑到 SCBS 之间还可以存在 D2D 通信, 使大规模无线网络中的那些处于单 SCBS 覆盖区域内的用户仍存在着关联 SCBS 还是接入 D2D 通信设备的接入选择问题。

#### 1) 基于时延最小化用户关联

文献[40]针对大规模无线网络中多 MEC 接入节点场景研究设计了基于时延最小化的用户关联方案。具体地, 文献[40]考虑一个多用户、多 MEC 节点密集部署的大规模无线网络。在这个网络中, 多个 SMD 可以通过无线信道卸载计算任务到其所关联的 MEC 服务器上。作者考虑了大规模无线网络中网络环境动态变化的特性, 例如时变信道条件、移动设备的可用能量以及不同 MEC 服务器的计算能力等, 针对 MEC 服务器的联合选择、协同卸载和切换问题, 提出了一种 MEC 服务器与集中式云相联合的协同计算卸载机制。在此基础上, 文献[40]提出了一种基于 DRL 的自适应用户关联、协作卸载和切换算法, 来实现最小化任务总计算时延。

基于时延最小化的用户关联算法虽然能够在很大程度上减少服务时延、提升用户 QoE, 但是这类关联策略通常没有考虑到本地设备能耗以及关联之后的传输能耗, 因此这类关联策略往往需要牺牲大量的系统能源来实现低时延目标。与之相反, 基于能耗最小化的用户关联策略则从能耗角度研究了多 MEC 节点系统下的用户接入选择问题。

#### 2) 基于能耗最小化用户关联

同样在 MEC 赋能的大规模无线网络中, 考虑 MEC 服务器的多接入情况, 文献[41]从 MEC 服务提供商的运营成本角度, 着重考虑了以降低计算卸载能耗为目标的联合用户接入选择以及用户应用迁移问题。作者针对用户关联问题, 考虑了大规模无线网络中的用户移动性, 提出了一种计算强度 (CI, computational intensity) 指标作为用户接入选择的准则。该指标根据应用的计算和数据卸载的需求来描述用户应用程序的类型, 然后根据应用类型来选择合适的 MEC 服务器进行关联, 同时选择在合适的 MEC 服务器上部署合适的应用服务内容。作者在上述关联准则的基础上, 将 MEC 服务器选择和用户应用迁移联合考虑, 将时延约束下的能量效率最大化问题建模为一个最短路径问题。实验结果表明, 随着 CI 的增加, MEC 服务器的选择由于能源效率而趋向于中心部署的关联方式, 然后又因为用户的容忍服务时延限制而返回到多 MEC 服务器关联部署的模式。同时, 文献[41]还发现, 在联合高精度的用户移动性预测以及可用资源预测下, 可以实现预先计算移动用户的关联选择策略以及预先在 MEC 服务器上进行应用部署, 以提高系统的

能量利用效率。

通过对上述用户关联问题现有研究工作的分析可以发现, 基于时延最小化的用户关联是一种以用户为中心的关联方式, 是为了提升用户的服务体验; 而基于能耗最小化的用户关联则是一种以网络为中心的关联方式, 是为了提高系统的能效, 减少运营成本。在实际系统中, 时延与能效通常需要联合考虑, 以实现最小能耗下的最优服务质量。

### 3) 基于能耗-时延权衡用户关联

文献[42-44]都从能耗和时延共同定义的用户 QoE 角度出发, 研究了能耗与时延权衡下的用户关联策略问题。

文献[42]研究了在 MEC 赋能的大规模无线网络中, 多 MEC 服务器的多任务卸载模型, 并对用户的关联问题和接入节点的业务部署问题进行了联合优化。考虑到任务卸载和执行所导致的时延增加和能量消耗, 文献[42]设计了一个基于时延和能量消耗的效用函数, 并以最小化该效用函数为目标来解决该多 MEC 服务器模型中的用户关联和业务部署问题。考虑到建模问题的复杂性, 作者设计了一种新的基于粒子群算法的优化算法来解决这一问题。文献[43]同样也从能耗和服务时延的角度出发, 定义了 QoE 感知的代价函数来作为网络中的性能指标, 通过对用户关联策略的优化来提升网络的系统性能。其中考虑的能耗产生于用户设备上行传输时的传输耗能, 服务时延定义为 SMD 到 MEC 服务器的传输时延和用户业务在 MEC 服务器上进行计算的计算时延之和。作者通过将上述用户接入选择问题建模成一个 0-1 整数规划问题, 提出了一种迭代优化求解算法, 来最小化其定义的 QoE 函数。文献[44]也是针对 MEC 赋能大规模无线网络中的用户接入选择问题, 提出了一种考虑用户 QoE 的用户关联选择算法来最大化用户的 QoE。在该研究工作中, 作者创建了统一的基于平均意见评分 (MOS, mean opinion score) 的 QoE 函数作为性能指标。MOS 的评分标准采用基于时延和能耗的伪主观评价方法, 在关注用户体验的同时, 定义满足接入节点选择实时性要求的 QoE 函数。最后, 文献[44]采用基于模拟退火算法的优化算法来求解该问题, 达到了比最大信噪比关联算法更优异的性能。

通过文献[42-44]的研究工作可以发现, 它们在进行能耗与时延联合优化下的用户关联问题研究时, 都是采用的加权和的方式来定义一个联合优化

目标, 这也是联合优化问题中的一种常见手段。然而通过本文前面的分析可以发现, 在 MEC 赋能的大规模无线网络中, 时延与能耗之间往往是具有复杂的耦合关系的, 而不是简单的线性加权关系。针对 MEC 赋能大规模无线网络基于更有效的能耗与时延联合优化用户关联问题, 博弈论<sup>[45]</sup>、机器学习<sup>[46]</sup>和多属性效用函数<sup>[47]</sup>等方法作为可供切入的方法, 需要被进一步研究。

### 3.5 隐私保护问题

在大规模无线网络中, 由于大量的 SCBS 的部署, 边缘网络中海量的接入用户以及繁多的应用业务, 使整个网络中的网络环境极其复杂。复杂的网络环境使黑客以及恶意用户、恶意无线节点的混入更容易, 边缘网络环境的安全性和隐私性受到极大威胁。此外, 在大规模无线网络中采用 MEC 技术之后, 由于边缘计算的任务卸载、内容感知、并行处理等开放特性, 使在密集接入的大规模无线网络环境下就已经存在的隐私保护问题得到进一步凸显<sup>[48]</sup>。具体地, MEC 赋能的大规模无线网络作为一个开放式的通信生态系统, 用户通常会授权关联的 MEC 节点对自己的数据进行处理与分析。在这一过程中, 若是某些 MEC 服务器被恶意篡改或操纵, 那么用户的隐私数据, 包括位置信息、身份信息、业务浏览记录等, 将会存在被窃取和泄露的风险。此外, 当用户中混入窃听者或者其他恶意用户时, 它们同样可以利用开放式的 MEC 系统, 通过 D2D 通信方式直接窃取其他用户的隐私数据, 或者通过访问 MEC 服务器节点来间接窃取服务器上存储的用户数据, 以实现非法盈利等目的。

MEC 赋能的大规模无线网络中的隐私数据泄露, 例如自动驾驶业务的用户数据泄露、医疗服务的用户数据泄露等, 将会对用户的人生财产安全造成极大的威胁, 同时也可能对服务提供商造成巨大的经济损失。因此, MEC 赋能的大规模无线网络中, 隐私保护问题具有重要的研究意义。目前, 该方面的研究工作主要可以从隐私类型进行分类, 分为用户位置隐私保护和用户数据隐私保护。本节将从这 2 个方面来对最近几年的隐私保护重点研究成果进行综述。

#### 1) 用户位置隐私保护

低时延与位置感知作为 MEC 系统的两大特点, 极大地推动了基于位置服务 (LBS, location-based service) 的应用, 扩展了 MEC 技术的应用范围<sup>[49]</sup>。

随着 LBS 的大量兴起,例如地图导航服务、位置共享服务等,在方便用户生活的同时也引起了人们对位置隐私保护的关注。相应地,MEC 系统下位置隐私保护机制(LPPM, location privacy preservation mechanism)是一个重要的研究课题,也有大量的工作进行这一方面的研究。

文献[50]研究了 MEC 系统中基于用户位置的服务迁移问题,其中服务迁移是指随着用户位置的不断变化,系统控制中心尽量把与用户相对应的服务内容布置到与用户靠近的 MEC 节点上,以进一步减少服务时延。作者将系统的总代价定义为服务迁移代价、服务时延和位置隐私泄露风险的组合。针对隐私风险的数学定义,文中考虑到业务需要迁移部署到用户周围的 MEC 节点上,因此,用户的位置离服务部署的距离越远,用户位置隐私泄露的风险越低,但是所带来的代价就是服务的时延越高。作者将系统总代价最小化问题建模成一个 Markov 决策过程,并提出了一种改进的策略迭代算法来求解最优业务迁移决策问题。与文献[50]考虑用户位置隐私保护与系统性能之间的权衡不同,文献[51]则专注于 MEC 系统中的位置隐私保护问题,致力于提升用户位置信息的抗窃取能力。具体地,文献[51]中假设系统中有窃听者的存在,考虑窃听者通过观察 MEC 系统中服务的迁移情况,来估计相应用户的位置轨迹信息。针对这种窃听者,为了保护用户的位置隐私,作者提出了一套基于 chaff 的保护策略来对抗窃听者的窃听:模仿策略,模仿用户的移动性;最大似然策略,最大化运动轨迹的似然概率;最优离线策略,根据用户的整个轨迹,最大限度地降低窃听者的跟踪精度;最优在线策略,基于用户历史轨迹,最小化期望跟踪精度。文中的研究与分析表明,在假设窃听者执行最大似然检测来估计用户的位置轨迹时,执行模范策略和最大似然策略作为对抗策略,可以将窃听者对用户的轨迹估计准确度降低到 10%~20%。同时,作者还在文中证明,当用户的移动是充分随机时,基于最优策略(离线或在线)的对抗策略可以将窃听者的跟踪精度衰减为 0。另外,文献[52]还从用户社交信息的角度,通过图学习的方法来学习用户的社交数据以满足基于位置的业务服务,避免了对用户位置信息的直接获取,降低了用户位置隐私泄露的风险。文献[52]所提思路虽然能够起到保护用户位置隐私的作用,但是用户的社交数据作为用户的一种

敏感数据,同样也需要得到保护。

## 2) 用户数据隐私保护

用户的隐私数据包括用户服务请求数据(如浏览记录)、本地存储的隐私信息(如银行卡数据)、用户社交数据(如家庭信息)等涉及用户隐私的各方面,是用户不愿公开的数据。用户隐私数据的泄露可能会严重威胁用户的人身财产安全。然而,很多以用户为中心的服务,例如基于用户兴趣的内容缓存、基于社交信息的推荐系统等,往往需要基于对用户数据的分析来保障业务的正常进行。因此,在保证用户隐私不被泄露的同时,允许对用户的隐私数据进行各类分析和操作是当前用户数据隐私保护的研究重点。

文献[53]研究了在具有动态网络环境的 MEC 系统中,如何在提供内容缓存业务的同时保护用户的隐私缓存数据。作者首先提出了一个基于区块链授权的分布式内容安全缓存框架,其中 SMD 执行内容缓存,MEC 服务器维护用于进行许可授权的区块链密钥,这样当有外界用户想要访问用户本地设备上的缓存内容时,必须要得到相应 MEC 服务器的访问授权才允许访问,这样就达到了保护用户本地隐私数据安全的目的。然后利用一种改进的 DRL 方法设计了一种考虑用户移动性的最佳内容缓存方案,以实现动态网络环境的自适应。最后考虑到传统区块链授权认证方法的复杂性,提出了一种新的区块链验证器选择方法来简化授权认证过程。通过将区块链技术和 DRL 技术相结合,实现了在保护用户缓存数据隐私的同时提供安全的内容缓存服务。文献[53]中用户相当于缓存内容的提供者,而其中用来保护用户隐私的区块链技术实际上相当于一种加密技术,通过密钥的认证授权来保证用户缓存数据的安全访问。

类似地,文献[54]考虑了 MEC 服务器作为缓存内容的提供者来提供缓存服务的场景,实际中的网盘服务就是这类场景的一个典型应用。用户通过将自己的一些数据上传保存到 MEC 服务中,以节省本地 SMD 的缓存资源。但是,如何保证用户数据在 MEC 服务器中的安全不泄露,关系到这类业务的用户数据隐私保护,是文献[54]的研究重点。在该文中,作者提出了一种基于博弈论的安全缓存方案,来保护用户的数据隐私。首先,为了保证提供安全的内容缓存业务,作者设计了一种基于直接信任度和间接信任度来推断 SCBS 信任度的信任评估

机制。根据信任度, 移动用户可以初步评估 SCBS 的安全性。接着, 为了在保护用户隐私的同时保证缓存内容的完整性, 提出了一种基于中国余数定理 (CRT, Chinese remainder theorem) 的内容加密协议。该协议包括片段加密和片段认证两部分。每个片段都用 CRT 加密成一个数字, 并且使用循环哈希链进行身份验证。最后通过 Stackelberg 博弈来研究 SMD 与 SCBS 之间的交互, 其中可信任的 SCBS 将会被选取用来为用户提供多归属访问的缓存业务, 并通过求解 SMD 与 SCBS 之间的联合平均利润共同最大化问题来实现 Stackelberg 均衡。

通过文献[53-54]可以发现, 它们都是从加密的角度来为用户的数据隐私提供保护。事实上, 在 MEC 赋能的大规模无线网络中的用户数据隐私保护的研究中, 除了文献[53]中基于区块链的加密机制以及文献[54]中基于博弈论的加密机制之外, 还有众多加密机制被用于用户数据隐私保护。例如文献[55]提出了基于概率公钥加密技术的来实现数据隐私保护的用户排名查询业务, 文献[56]提出了基于混沌系统伪随机置换的加密技术来保护用户数据隐私。值得注意的是, 虽然这些基于加密机制的隐私保护方案能够在提供应用服务的同时在一定程度上保护用户的隐私数据, 但是这类方法都必须引入额外的加密、认证、解密操作, 这无疑会增加系统额外的资源消耗, 降低系统的资源利用效率。因此, 研究解决用户数据隐私保护与服务质量之间的矛盾是未来的一个重要研究方向。

### 3.6 总结与展望

前文从计算卸载问题、边缘缓存问题、多维资源分配问题、用户关联问题和隐私保护问题这 5 个角度出发, 综述了移动边缘计算与缓存技术赋能的大规模无线网络中仍亟待解决的关键问题以及现有的一些解决方案。表 1 从关键问题、关键研究点、方法论 3 个方面对上述解决方案进行了综合分析 with 总结。

当前研究工作中, 对于边缘网络中通信、计算、缓存资源的考虑相对独立, 而新型应用的实时性需求一般较高, 需要深度融合这三者以降低时延和网络负载; 需要深入研究编码缓存与边缘计算之间的交互机理和边缘网络编码缓存方案的自动化设计与自适应调整; 探索建立能定量表征计算、缓存资源转化为通信增益的多维混合资源调配优化模型, 打破传统模型中通信、计算、缓存资源相对独立的

架构, 实现资源有机整合。

未来的研究方向还包括以下几个方面。

1) 考虑卸载、缓存、资源分配和用户关联相混合的多维度联合优化, 系统性地研究各个维度的策略对系统性能的影响, 揭示 MEC 赋能的大规模无线网络中这些关键问题之间的内在关系。

2) 关注用户隐私保护, 研究隐私保护与系统性能之间的矛盾, 探索 MEC 赋能大规模无线网络架构下新型的隐私保护方案。

3) 探索人工智能驱动的边缘密集网络, 研究如深度学习、强化学习、联邦学习等在内的机器学习方法对解决上述关键各类问题所能提供的帮助。

4) 在缓存使能的大规模网络中使用数据驱动的深度学习与无线传输物理层理论, 设计自洽式的传输方案。

5) 构建去中心化的云边协同以及各边缘节点间协同传输机制, 挖掘边缘网络中边缘缓存的容量性能极限。

6) 研究大规模网络中无线缓存网络的物理层传输机制, 实现低时延、高吞吐率的边缘网络通信。为无线缓存网络边缘节点自洽式智能干扰管理建立数学模型, 以解决由于大量信息交互引起的时延开销增大以及通信质量下降等问题。

7) 为支持新型边缘计算任务, 设计新型无线缓存方案。例如, 为边缘计算场景下的联邦学习任务提供无线缓存服务, 揭示联邦学习任务的收敛速度和性能。

## 4 隐私保护的边缘计算与缓存

如图 3 所示, 在典型 MEC 赋能的大规模无线网络中, 可以通过深度卷积神经网络来提供一种图像分类业务。用户可以通过 MEC 节点获取分类器来对自己所需要分类的图片进行分类, 从而实现 MEC 节点对该分类业务的交付。具体地, 主要有 2 种实现方式。第一, 假设网络中某些用户作为服务的请求者, 向附近的 MEC 接入节点请求对自己本地设备上的某一张或某几张图片进行分类。此时, 收到分类业务请求的 MEC 节点, 就会利用存储在其 MEC 服务器中的图片分类器对用户的待分类图片进行分类, 再返回用户相应的分类结果。第二, 收到分类业务请求的 MEC 节点, 直接将存储在其 MEC 服务器中的图片分类器下发给对应用户, 让其自行执行分类操作。可见, 分类器是该类业务的核心

表 1 MEC 赋能大规模网络中关键问题和解决方案的分析与总结

关键问题	关键研究点	相关文献	方法论
计算卸载问题	能耗敏感型计算卸载	文献[15-16]	整体卸载，基于 SCA 的迭代优化的卸载 部分卸载，基于阈值优先级的计算卸载
	时延敏感型计算卸载	文献[17]	基于 Lyapunov 的低复杂度在线动态计算卸载
	时延-能耗敏感型计算卸载	文献[18-19]	基于惩罚函数和 D.C.规划的迭代搜索计算卸载 基于交替凸搜索的计算卸载
边缘缓存问题	非编码缓存	文献[21-26]	基于统计用户请求的历史数据 基于优化算法 基于动态规划和强化学习
	编码缓存	文献[20,27-30]	基于深度神经网络的 PDA 设计 基于强化学习的缓存更新和多播调度 基于优化算法
多维资源分配问题	计算资源分配	文献[31-32]	时间维度的资源分配，基于 Lyapunov 的最优计算资源分配 空间维度的资源分配，基于势博弈论的计算资源分配
	缓存资源分配	文献[33-34]	基于用户社交关系的静态缓存资源配置 基于用户时变内容请求的在线响应式动态缓存资源分配
	能量资源分配	文献[35-36]	CSI 和 TSI 知识完备时，基于凸优化的离线传输功率分配算法实；CSI 和 TSI 知识仅因果已知，基于启发式的在线传输功率分配 基于拉格朗日对偶法的离线能耗优化算法；基于滑动窗口以及序列优化的在线传输功率分配算法
	通信资源分配	文献[37]	基于回声状态网络和迁移学习的上下行传输资源块动态分配
	混合资源分配	文献[38-39]	基于深度强化学习的通信与计算资源联合分配 基于自适应优化算法、粒子群算法与 Pareto 一对一匹配算法相结合的计算、能量和通信资源混合资源分配
用户关联问题	基于时延最小化用户关联	文献[40]	基于深度强化学习的自适应用户关联算法，实现任务总计算时延最小化
	基于能耗最小化用户关联	文献[41]	基于最短路径优化的用户关联算法，实现时延约束下的能效最大化
	基于能耗-时延权衡用户关联	文献[42-44]	基于粒子群的用户关联，实现时延能耗联合代价最小化 基于 0-1 整数规划的迭代优化用户关联，实现时延能耗联合定义的 QoE 代价最小化 基于模拟退火的用户关联，实现用户 QoE 最大化
隐私保护问题	用户位置隐私保护	文献[50-52]	基于 Markov 决策的策略迭代算法来保护用户位置隐私的同时实现最优业务迁移决策 基于 chaff 的保护策略来对抗窃听器，保护用户位置隐私 基于图学习来保护用户位置隐私的同时实现基本的业务提供
	用户数据隐私保护	文献[53-56]	深度强化学习联合区块链授权与认证的用户数据加密技术，实现安全的内容缓存服务 博弈论联合中国余数论的内容加密协议，实现用户数据隐私保护下的多归属访问缓存业务 基于概率论的公钥加密技术，实现数据隐私保护下的用户排名查询业务 基于混沌系统的伪随机置换加密技术

心。鉴于基于深度卷积神经网络的分类器是处理当前业务最常见和高效的手段，故本文以此来实现该类图像分类业务。需要指出的是，这类分类器往往需要经过大量的数据训练过程之后，才能达到足够高的分类准确率。

在该通信网络架构下，集中式的训练是一种常用的训练方法。具体地，假设每个用户在本地设备上都拥有一定数量的标签图像数据，但是受限于存

储能力以及用户习惯，每个用户设备上的标签图像数据数量有限，无法单独支撑起一个分类器的训练。为此，MEC 服务器可以通过收集所有用户的数据，在 MEC 服务器端得到一个总的数据集。接着，利用其强大的计算能力，在服务器端对分类器进行集中式训练。文献[57]表明，这种集中式的训练方法可以使分类器达到非常高的准确率。但是，由于该集中式的训练过程涉及用户本地的图像数据的

收集, 可能对用户的数据隐私安全造成极大的威胁。而在隐私保护日益受到重视的当今社会, 上述威胁用户隐私的集中式训练方式是无法被容忍的。鉴于此, 本文充分利用 MEC 赋能大规模无线网络中的丰富计算资源与缓存资源, 提出了一种基于联邦学习<sup>[58]</sup>的分布式训练架构, 在对分类器进行有效训练的同时, 还实现对用户数据隐私的保护。此外, 本文还利用 MEC 服务器中丰富的缓存资源对传统的联邦学习方案进行了改进, 提出了一种带缓存的联邦学习算法。该算法能够在保护用户数据隐私的同时, 提高分类器的分类准确率, 从而有效改善用户的服务质量。

#### 4.1 隐私保护方案

在存在  $I$  个 SMD、 $S$  个 SCBS 和一个 MBS 的 MEC 赋能大规模无线网络场景中, 本文利用 SMD 与 MEC 服务器都具有计算能力的特点, 提出了一种基于联邦学习的训练框架。具体地, 假设每个 SMD 上都能部署一个分类器网络, 第  $i$  个 SMD 上部署的分类器网络参数可以表示为  $\theta^i$ , 每个用户都可以在本地利用自身的图片数据集单独训练自己的分类器网络。不过, 如果用户仅依靠本地的少量数据集来对分类器进行训练, 分类器将难以达到一个较高分类准确率。因此, 基于联邦学习的框架使每个 SCBS 每隔一定的时间并行且同步地从各自服务范围内随机挑选一个用户, 表示为  $a_{i_s}^s = 1, \forall i_s \in \mathcal{I}_s$ 。其中,  $i_s$  为第  $s$  个 SCBS 所挑选到的用户编号,  $\mathcal{I}_s$  为第  $s$  个 SCBS 服务范围内的用户集合。当用户  $i_s \in \mathcal{I}_s$  未被挑选时,  $a_{i_s}^s = 0$ 。在  $s$  号 SCBS 服务区内, 当  $a_{i_s}^s = 1$  时, 该 SCBS 将发送自己当前所拥有的全局模型  $\theta^G(t)$  给用户  $i_s$  进行  $E$  次本地训练。本地训练结束后, 用户  $i_s$  上传训练后的本地分类器的网络参数  $\theta^{i_s}(t+1)$  给其所连接的 SCBS。每个 SCBS 接收到上传的网络参数后, 通过回程链路上传给 MBS 进行参数聚合, 即  $\theta^G(t+1) = \frac{1}{S} \sum_{s=1}^S \theta^{i_s}(t+1)$ 。聚合之后, MBS 将  $\theta^G(t+1)$  发送给每个 SCBS。接着, 每个 SCBS 将再次并行且同步地从各自服务范围内随机挑选一个用户, 然后重复上述通信交互过程。

一般地, 上述通信交互过程可被称为一个通信回合, 其中  $t$  代表着第  $t$  次通信回合。每个通信回合中得到的聚合参数体现了本轮所有参与聚合的

用户的本地数据特征, 可见其间接地扩充了分类器的训练数据。同时, 在每个通信回合中, SMD 与 SCBS 之间仅进行了分类器网络参数的交互, 而不涉及任何与用户本地图片数据有关的交互, 因此有效地保护了用户的隐私。

此外, 针对上述基于联邦学习的分布式训练方案, 利用 MEC 服务器上的缓存资源, 本文对上述方案进行了改进, 提出了一种带缓存的联邦学习训练方案。具体而言, 在第  $t$  个通信回合, 每个 SCBS 收到  $\theta^G(t)$  之后, 将其存储到自己的 MEC 服务器中。在下一通信回合到来后, 每个 SCBS 将再从 MBS 收到该回合的聚合参数  $\theta^G(t+1)$ 。此时, 每个 SCBS 将对第  $t$  次和第  $t+1$  次的聚合参数进行一个加权和以得到新的聚合参数, 即  $\theta^G(t+1) \leftarrow \lambda \theta^G(t) + (1-\lambda) \theta^G(t+1)$ , 其中  $\lambda$  为加权因子。接着, 每个 SCBS 再将新参数  $\theta^G(t+1)$  广播给各自服务范围内的用户, 并同时存储到自己的 MEC 服务器中, 以用于进行下一个回合的加权求和操作。

上述训练期间的参数传递流程可用图 5 表示。在整个训练期间, 通信系统中不涉及任何与用户隐私有关的数据传递。在该联邦学习框架中, 仅有模型参数在进行交互与传递, 因此该过程不会涉及。该考虑隐私保护的带缓存的联邦学习训练算法如算法 1 所示。

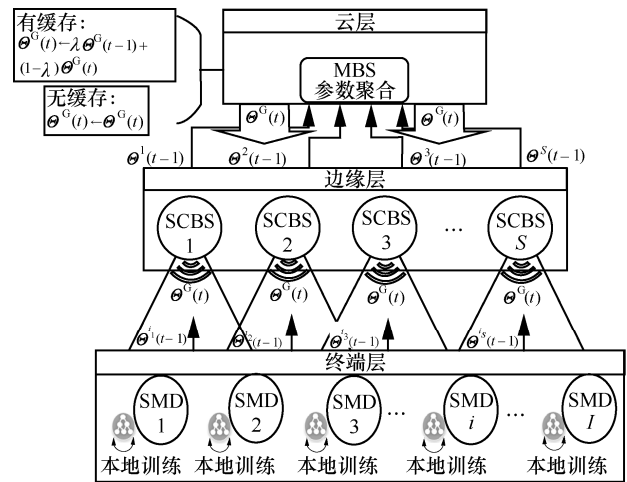


图 5 隐私保护的带缓存的联邦学习训练算法参数传递流程

**算法 1** 隐私保护的带缓存联邦学习训练算法  
MBS 端随机初始化  $\theta^G(0)$ ;

对每个 SCBS,  $\forall s \in \{1, 2, \dots, S\}$  并行初始化指令参数  $a_{i_s}^s = 0, \forall i_s \in \mathcal{I}_s$ ; 初始化回合数  $t=0$ ;

- 1) 循环
- 2) MBS 发送  $\theta^G(t)$  给所有的 SCBS;
- 3) 对  $\forall s \in \{1, 2, \dots, S\}$ , 每个 SCBS 并行且同步执行:
  - 4) 随机挑选一个用户  $i_s \in \mathcal{I}_s$ , 向其发送指令  $a_{i_s}^s = 1$  和参数  $\theta^G(t)$ ;
  - 5) 每个 SCBS 服务区内被挑选的用户并行执行:
    - 6)  $\theta^{i_s}(t) \leftarrow \theta^G(t)$ ;
    - 7) 进行  $E$  次本地训练;
    - 8) 得到训练后的  $\theta^{i_s}(t+1)$  并上传
    - 9) 每个 SCBS 接收到各自服务区内用户上传的  $\theta^{i_s}(t+1)$ , 并同时上传给 MBS;
  - 10) MBS 聚合  $\theta^G(t+1) = \frac{1}{S} \sum_{s=1}^S \theta^{i_s}(t+1)$ ;
  - 11)  $\theta^G(t+1) \leftarrow \lambda \theta^G(t) + (1-\lambda) \theta^G(t+1)$ ;
  - 12) MBS 删除  $\theta^G(t)$ , 且缓存  $\theta^G(t+1)$ ;
  - 13)  $t = t + 1$ ;
  - 14) until 算法满足收敛条件, 即模型的训练准确率在前后连续 2 个回合内的变化误差小于阈值  $\varepsilon$ ;
  - 15) return  $\theta^G(t+1)$ ;
  - 16) MBS 通过各个 SCBS 将  $\theta^G(t+1)$  发送给所有用户。

#### 4.2 实验与仿真

典型的图像分类任务如下。假设每个用户都拥有一定数量的标签图像数据, 这些数据来源于 Cifar-10<sup>[57]</sup>数据集, 并且每个用户训练数据的类别分布  $q$  服从参数为  $\alpha$  的迪利克雷分布  $\text{Dir}(\alpha)$ 。为了保证用户之间数据分布呈现非独立同分布特性, 本文取  $\alpha = 0.5$ 。假设每个用户都可以在本地使用相同的神经网络进行模型的训练, 且分类器模型都为 ResNet18<sup>[59]</sup>。

图 6 和图 7 分别给出了总用户数为 400 和 500 时的算法性能对比仿真结果, 其中 SCBS 数目为 10,  $\lambda = 0.990$ 。从图 6 中可以看出, 无论是否使用缓存策略, FL 算法都可以收敛, 但是采用缓存策略的 FL 算法取得了更高的测试准确率。结合图 7 和表 2 可以发现, 当  $\lambda = 0.990$ 、用户数为 500 时, 采用缓存策略后仅提升了 0.34% 的性能。该结果表明具有缓存策略的 FL 可以在一定程度上抵制用户数目增大带来的泛化性能损失。然而, 缓存所带来的增益

同时受到多个参数影响, 如每次通信处于激活状态的接入点数量和本地训练所使用模型的容量等, 这导致了随着用户数目的增大, 缓存的增益趋于递减的趋势。不过, 需要指出的是, 进一步调整缓存权重因子  $\lambda$ , 所提算法的性能能够获得进一步提升。例如, 通过表 2 可以发现, 当总用户数为 500 时, 设置  $\lambda = 0.991$ , 所提算法将能够达到 1.48% 的识别准确率提升。同样地, 当用户数为 400 时, 设置  $\lambda = 0.997$ , 所提算法可以获得 2.33% 的准确率提升。

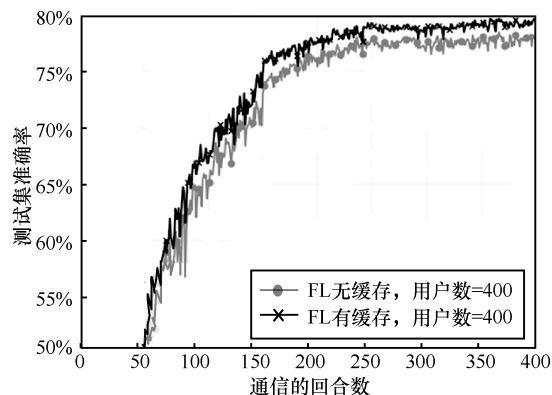


图 6 总用户数为 400 时 2 种方案训练过程对比 ( $\lambda = 0.990$ )

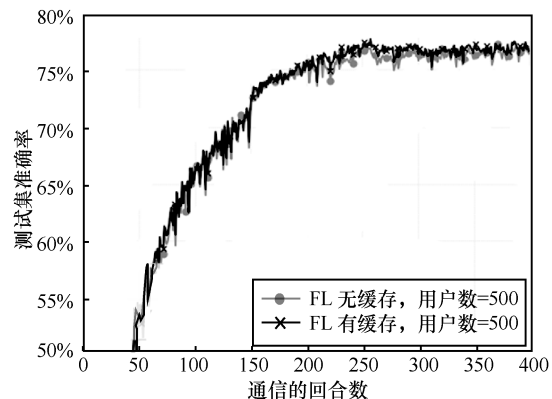


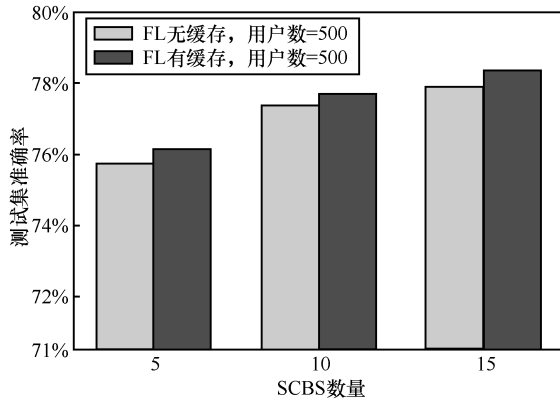
图 7 总用户数为 500 时 2 种方案训练过程对比 ( $\lambda = 0.990$ )

此外, 从表 2 中还可以看出, 随着用户数目增大, 2 种方案的准确率都有所下降, 这主要是因为用户数目增大导致每个用户分得的本地数据减小, 本地训练过拟合的风险增加。

图 8 给出了在不同 SCBS 数量下, 2 种方案在分类准确率上的对比。从图 8 中可以看出, 在固定用户数为 500 的场景中, 随着通信接入点数量的增大, FL 的性能有不同程度的提升。如通信接入点为 5、10 和 15 时, 所提算法可以一直获得比传统无缓存的 FL 方法更高的准确率。

表 2 2 种方案在不同用户数上的性能对比

方案	总用户数	测试集准确率									
		$\lambda=0.990$	$\lambda=0.991$	$\lambda=0.992$	$\lambda=0.993$	$\lambda=0.994$	$\lambda=0.995$	$\lambda=0.996$	$\lambda=0.997$	$\lambda=0.998$	$\lambda=0.999$
FL 无缓存	400	78.47%									
FL 有缓存		79.68%	79.05%	79.19%	80.40%	79.62%	80.78%	79.7%	<b>80.80%</b>	80.11%	78.84%
FL 无缓存	500	77.36%									
FL 有缓存		77.70%	<b>78.84%</b>	77.64%	76.65%	77.77%	78.05%	77.92%	78.70%	77.30%	78.03%

图 8 2 种方案在不同 SCBS 数量上的性能对比 ( $\lambda=0.990$ )

综上所述, 在边缘计算任务中, 所提算法可以在保证不侵犯用户数据隐私的条件下使用 FL 框架, 充分利用分散在用户本地设备的数据和算力, 以获得整个计算任务的性能提升。同时, 为联邦学习框架增加缓存策略, 通过合理利用联邦学习优化过程中产生的历史信息, 可以进一步提升边缘计算任务的收敛速度和性能。

## 5 结束语

随着信息化时代的到来, 各类新型无线业务呈现出爆发式增长, 对无线网络提出了严峻挑战。移动边缘计算与缓存技术赋能大规模无线网络能够同时满足大规模接入、密集计算与密集缓存等业务需求, 被作为一项非常具有前景的技术用于解决上述挑战。本文分别从大规模无线网络中的移动边缘计算与移动边缘缓存的基本概念出发, 介绍了三者的相关背景知识。接着, 针对未来大规模无线网络的移动边缘计算与缓存技术, 综述了最近研究工作在计算卸载问题、边缘缓存问题、多维资源分配问题、用户关联问题和隐私保护问题这些关键问题上的相关研究, 并指出了未来的一些研究重点和研究方向。最后, 本文针对关键问题中的隐私保护问题, 提出了一种 MEC 赋能大规模无线网络中基

于联邦学习的隐私保护的业务服务架构, 并进一步利用 MEC 服务器上的丰富缓存资源提出了一种带缓存的改进型联邦学习算法, 实现了保护用户隐私的同时, 提升服务的质量。

## 参考文献:

- [1] YOU X H, WANG C X, HUANG J, et al. Towards 6G wireless communication networks: vision, enabling technologies, and new paradigm shifts[J]. Science China Information Sciences, 2020, 64(1): 1-74.
- [2] CISCO. Cisco visual networking index: global mobile data traffic forecast update, 2017-2022[R]. Cisco Public Information, 2017.
- [3] HE S W, HUANG W, WANG J H, et al. Cache-enabled coordinated mobile edge network: opportunities and challenges[J]. IEEE Wireless Communications, 2020, 27(2): 204-211.
- [4] GUO H Z, LIU J J, ZHANG J. Computation offloading for multi-access mobile edge computing in ultra-dense networks[J]. IEEE Communications Magazine, 2018, 56(8): 14-19.
- [5] KAMEL M, HAMOUDA W, YOUSSEF A. Ultra-dense networks: a survey[J]. IEEE Communications Surveys & Tutorials, 2016, 18(4): 2522-2545.
- [6] ABBAS N, ZHANG Y, TAHERKORDI A, et al. Mobile edge computing: a survey[J]. IEEE Internet of Things Journal, 2018, 5(1): 450-465.
- [7] 谢人超, 廉晓飞, 贾庆民, 等. 移动边缘计算卸载技术综述[J]. 通信学报, 2018, 39(11): 138-155.
- [8] XIE R C, LIAN X F, JIA Q M, et al. Survey on computation offloading in mobile edge computing[J]. Journal on Communications, 2018, 39(11): 138-155.
- [9] ZHANG K, LENG S P, HE Y J, et al. Mobile edge computing and networking for green and low-latency Internet of things[J]. IEEE Communications Magazine, 2018, 56(5): 39-45.
- [10] PAN C H, ELKASHLAN M, WANG J Z, et al. User-centric C-RAN architecture for ultra-dense 5G networks: challenges and methodologies[J]. IEEE Communications Magazine, 2018, 56(6): 14-20.
- [11] TIAN X J, JIA W J. Improved clustering and resource allocation for ultra-dense networks[J]. China Communications, 2020, 17(2): 220-231.
- [12] ETSI. ETSI GS MEC 002. Mobile edge computing (MEC); technical requirements[R]. ETSI White Paper, 2016.
- [12] 代玥玥. 移动边缘计算中资源管理问题的研究[D]. 成都: 电子科技大学, 2019.
- DAI Y Y. Research on resource management problems in mobile edge

- computing[D]. Chengdu: University of Electronic Science and Technology of China, 2019.
- [13] ZENG J, SUN J Y, WU B W, et al. Mobile edge communications, computing, and caching (MEC3) technology in the maritime communication network[J]. *China Communications*, 2020, 17(5): 223-234.
- [14] HE Y, YU F R, ZHAO N, et al. Secure social networks in 5G systems with mobile edge computing, caching, and device-to-device communications[J]. *IEEE Wireless Communications*, 2018, 25(3): 103-109.
- [15] HABER E E, NGUYEN T M, ASSI C, et al. An energy-efficient task offloading solution for MEC-based IoT in Ultra-dense networks[C]//2019 IEEE Wireless Communications and Networking Conference. Piscataway: IEEE Press, 2019: 1-7.
- [16] YOU C S, HUANG K B. Multiuser resource allocation for mobile-edge computation offloading[C]//2016 IEEE Global Communications Conference. Piscataway: IEEE Press, 2016: 1-6.
- [17] MAO Y Y, ZHANG J, LETAIEF K B. Dynamic computation offloading for mobile-edge computing with energy harvesting devices[J]. *IEEE Journal on Selected Areas in Communications*, 2016, 34(12): 3590-3605.
- [18] ZHANG J, HU X P, NING Z L, et al. Energy-latency tradeoff for energy-aware offloading in mobile edge computing networks[J]. *IEEE Internet of Things Journal*, 2018, 5(4): 2633-2645.
- [19] QIN M, CHENG N, JING Z W, et al. Service-oriented energy-latency tradeoff for IoT task partial offloading in MEC-enhanced multi-RAT networks[J]. *IEEE Internet of Things Journal*, 2021, 8(3): 1896-1907.
- [20] MADDAH-ALI M A, NIESEN U. Fundamental limits of caching[J]. *IEEE Transactions on Information Theory*, 2014, 60(5): 1077-1081.
- [21] BORST S C, GUPTA V, WALID A. Distributed caching algorithms for content distribution networks[C]// IEEE Conference on Information Communications. Piscataway: IEEE Press, 2010: 1-9.
- [22] TAMOOR-UL-HASSAN S, BENNIS M, NARDELLI P H J, et al. Caching in wireless small cell networks: a storage-bandwidth tradeoff[J]. *IEEE Communications Letters*, 2016, 20(6): 1175-1178.
- [23] ZHU K L, ZHI W T, CHEN X, et al. Socially motivated data caching in ultra-dense small cell networks[J]. *IEEE Network*, 2017, 31(4): 42-48.
- [24] DAI C, ZHU K, WANG R, et al. Contextual multi-armed bandit for cache-aware decoupled multiple association in UDNs: a deep learning approach[J]. *IEEE Transactions on Cognitive Communications and Networking*, 2019, 5(4): 1046-1059.
- [25] TANG J, DAI T W, CUI M M, et al. Optimization for maximizing sum secrecy rate in SWIPT-enabled NOMA systems[J]. *IEEE Access*, 2018, 6: 43440-43449.
- [26] ZHANG Z M, YANG Y Q, HUA M, et al. Proactive caching for vehicular multi-view 3D video streaming via deep reinforcement learning[J]. *IEEE Transactions on Wireless Communications*, 2019, 18(5): 2693-2706.
- [27] KARAMCHANDANI N, NIESEN U, MADDAH-ALI M A, et al. Hierarchical coded caching[J]. *IEEE Transactions on Information Theory*, 2016, 62(6): 3212-3229.
- [28] JI M Y, CAIRE G, MOLISCH A F. Fundamental limits of caching in wireless D2D networks[J]. *IEEE Transactions on Information Theory*, 2016, 62(2): 849-869.
- [29] YAN Q F, CHENG M Q, TANG X H, et al. On the placement delivery array design for centralized coded caching scheme[J]. *IEEE Transactions on Information Theory*, 2017, 63(9): 5821-5833.
- [30] ZHANG Z M, CHEN H Y, HUA M, et al. Double coded caching in ultra dense networks: caching and multicast scheduling via deep reinforcement learning[J]. *IEEE Transactions on Communications*, 2020, 68(2): 1071-1086.
- [31] YANG X, CHEN Z Y, LI K K, et al. Communication-constrained mobile edge computing systems for wireless virtual reality: scheduling and tradeoff[J]. *IEEE Access*, 2018, 6: 16665-16677.
- [32] LIU H, JIA H M, CHEN J Q, et al. Computing resource allocation of mobile edge computing networks based on potential game theory[C]//2018 IEEE 4th International Conference on Computer and Communications. Piscataway: IEEE Press, 2018: 693-699.
- [33] NIKOLAOU S, RENESSE R V, SCHIPER N. Proactive cache placement on cooperative client caches for online social networks[J]. *IEEE Transactions on Parallel and Distributed Systems*, 2016, 27(4): 1174-1186.
- [34] AZIMI S M, SIMEONE O, SENGUPTA A, et al. Online edge caching and wireless delivery in fog-aided networks with dynamic content popularity[J]. *IEEE Journal on Selected Areas in Communications*, 2018, 36(6): 1189-1202.
- [35] WANG F, XU J, CUI S G. Optimal energy allocation and task offloading policy for wireless powered mobile edge computing systems[J]. *IEEE Transactions on Wireless Communications*, 2020, 19(4): 2443-2459.
- [36] WANG F, XING H, XU J. Real-time resource allocation for wireless powered multiuser mobile edge computing with energy and task causality[J]. *IEEE Transactions on Communications*, 2020, 68(11): 7140-7155.
- [37] CHEN M Z, SAAD W, YIN C C, et al. Data correlation-aware resource management in wireless virtual reality (VR): an echo state transfer learning approach[J]. *IEEE Transactions on Communications*, 2019, 67(6): 4267-4280.
- [38] WANG F, XING H, XU J. Real-time resource allocation for wireless powered multiuser mobile edge computing with energy and task causality[J]. *IEEE Transactions on Communications*, 2020, 68(11): 7140-7155.
- [39] DIAO X B, ZHENG J C, WU Y, et al. Joint computing resource, power, and channel allocations for D2D-assisted and NOMA-based mobile edge computing[J]. *IEEE Access*, 2019, 7: 9243-9257.
- [40] HO T M, NGUYEN K K. Joint server selection, cooperative offloading and handover in multi-access edge computing wireless network: A deep reinforcement learning approach[J]. *IEEE Transactions on Mobile Computing*, 2020, PP(99): 1536-1233.
- [41] THANANJEYAN S, CHAN C A, WONG E, et al. Mobility-aware energy optimization in hosts selection for computation offloading in multi-access edge computing[J]. *IEEE Open Journal of the Communications Society*, 2020, 1: 1056-1065.
- [42] MA S Y, SONG S D, ZHAO J M, et al. Joint network selection and service placement based on particle swarm optimization for multi-access edge computing[J]. *IEEE Access*, 2020, 8: 160871-160881.
- [43] ZHAO J M, WU W J, GUO X, et al. Access selection considering mobile edge computing in ultra dense network[C]//2017 3rd IEEE International Conference on Computer and Communications. Piscataway: IEEE Press, 2017: 433-437.
- [44] GAO Y, WU W J, ZHOU T Q, et al. QoE-aware access node selection considering mobile edge computing[C]//2018 IEEE 4th International Conference on Computer and Communications. Piscataway: IEEE

- Press, 2018: 1914-1918.
- [45] SAAD W, HAN Z, DEBBAH M, et al. Coalitional game theory for communication networks[J]. IEEE Signal Processing Magazine, 2009, 26(5): 77-97.
- [46] ZHENG C, LIU S H, HUANG Y M, et al. MEC-enabled wireless VR video service: a learning-based mixed strategy for energy-latency tradeoff[C]//2020 IEEE Wireless Communications and Networking Conference. Piscataway: IEEE Press, 2020: 1-6.
- [47] ABDILDIN Y G, ABBAS A E. An algorithm for excluding redundant assessments in a multiattribute utility problem[J]. Procedia Computer Science, 2012, 9(1): 802-811.
- [48] 张佳乐, 赵彦超, 陈兵, 等. 边缘计算数据安全与隐私保护研究综述[J]. 通信学报, 2018, 39(3): 1-21.  
ZHANG J L, ZHAO Y C, CHEN B, et al. Survey on data security and privacy-preserving for the research of edge computing[J]. Journal on Communications, 2018, 39(3): 1-21.
- [49] TIAN Z H, WANG Y H, SUN Y B, et al. Location privacy challenges in mobile edge computing: classification and exploration[J]. IEEE Network, 2020, 34(2): 52-56.
- [50] WANG W X, GE S X, ZHOU X B. Location-privacy-aware service migration in mobile edge computing[C]//2020 IEEE Wireless Communications and Networking Conference. Piscataway: IEEE Press, 2020: 1-6.
- [51] HE T, CIFTCIOGLU E N, WANG S Q, et al. Location privacy in mobile edge clouds: a chaff-based approach[J]. IEEE Journal on Selected Areas in Communications, 2017, 35(11): 2625-2636.
- [52] WU Q, CHEN X, ZHOU Z, et al. Mobile social data learning for user-centric location prediction with application in mobile edge service migration[J]. IEEE Internet of Things Journal, 2019, 6(5): 7737-7747.
- [53] DAI Y Y, XU D, ZHANG K, et al. Deep reinforcement learning and permissioned blockchain for content caching in vehicular edge computing and networks[J]. IEEE Transactions on Vehicular Technology, 2020, 69(4): 4312-4324.
- [54] XU Q C, SU Z, ZHENG Q H, et al. Game theoretical secure caching scheme in multihoming edge computing-enabled heterogeneous networks[J]. IEEE Internet of Things Journal, 2019, 6(3): 4536-4546.
- [55] PASUPULETI S K, RAMALINGAM S, BUYYA R. An efficient and secure privacy-preserving approach for outsourced data of resource constrained mobile devices in cloud computing[J]. Journal of Network and Computer Applications, 2016, 64: 12-22.
- [56] BAHRAMI M, SINGHAL M. A light-weight permutation based method for data privacy in mobile cloud computing[C]//2015 3rd IEEE International Conference on Mobile Cloud Computing, Services, and Engineering. Piscataway: IEEE Press, 2015: 189-198.
- [57] KRIZHEVSKY A, HINTON G. Learning multiple layers of features from tiny images[R]. Handbook of Systemic Autoimmune Diseases,

(2009-01-04)[2021-01-21].

- [58] KAMP M, ADILOVA L, SICKING J, et al. Efficient decentralized deep learning by dynamic model averaging[C]// Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Berlin: Springer, 2018: 393-409.
- [59] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 770-778.

#### [作者简介]



**黄永明**（1977- ），男，江苏吴江人，博士，东南大学教授、博士生导师，主要研究方向为智能 5G/6G 移动通信、毫米波无线通信等。



**郑冲**（1994- ），男，湖北荆州人，东南大学博士生，主要研究方向为智能无线通信、移动边缘计算、边缘智能、物联网、深度强化学习和联邦学习等。



**张征明**（1994- ），男，安徽阜阳人，东南大学博士生，主要研究方向为无线大数据、机器学习、5G 移动网络、无人机辅助通信和资源管理等。



**尤肖虎**（1962- ），男，山东济宁人，博士，东南大学教授、博士生导师，主要研究方向为移动通信系统和信号处理及其应用等。