

# 面向边缘计算的资源 优化技术研究进展

屈志昊<sup>1</sup>, 叶保留<sup>2</sup>, 陈贵海<sup>2</sup>, 唐斌<sup>2</sup>, 郭成昊<sup>3</sup>

1. 河海大学计算机与信息学院, 江苏 南京 211100; 2. 南京大学计算机科学与技术系, 江苏 南京 210046;

3. 中国电子科技集团公司第二十八研究所, 江苏 南京 210007

## 摘要

以云计算模型为代表的集中式数据处理关键技术已不能高效、及时地处理边缘设备产生的数据。针对这一问题, 以“数据处理应更靠近数据源头”为核心理念的边缘计算模型应运而生。首先介绍了微数据中心、微云、雾计算、移动边缘计算等计算范型, 并讨论了边缘资源整合的优势。然后, 回顾了近年来边缘计算中与资源优化领域相关的工作, 以计算、存储和通信3种资源为切入点, 分别从计算卸载、分布式缓存和高性能传输这3个研究热点, 对国内外的研究进展进行总结和讨论。最后, 展望了该领域未来的发展趋势和主要的研究方向。

## 关键词

边缘计算; 计算卸载; 分布式缓存; 高性能传输

中图分类号: TP393

文献标识码: A

doi: 10.11959/j.issn.2096-0271.2019011

## State-of-the-art survey on resource optimization in edge computing

QU Zhihao<sup>1</sup>, YE Baoliu<sup>2</sup>, CHEN Guihai<sup>2</sup>, TANG Bin<sup>2</sup>, GUO Chenghao<sup>3</sup>

1. College of Computer and Information, Hohai University, Nanjing 211100, China

2. Department of Computer Science and Technology, Nanjing University, Nanjing 210046, China

3. The 28th Research Institute of China Electronics Technology Group Corporation, Nanjing 2100007, China

## Abstract

The traditional centralized architecture, known as cloud computing, cannot accommodate such user demands in an efficient and timely manner. To cope with this problem, edge computing architectures have been proposed with the core concept of that “data processing should be close to the data source”. Firstly, paradigms of edge computing was introduced, including micro data center, cloudlet, fog computing, and mobile edge computing, and the advantages of edge computing from the perspective of resource integration was discussed. Then, related works of resource optimization in edge computing was reviewed and summarized, and these works was discussed via three directions, i.e., computation offloading, distributed caching and high performance transmission, corresponding to core resources as computing, storage and communication. Finally, trends of development and future directions were presented as well.

## Key words

edge computing, computation offloading, distributed caching, high performance transmission

## 1 引言

近年来,随着信息技术的不断进步以及智能终端设备的不断泛在普及,基于移动互联网的智能化应用蓬勃发展,并在人们日常生活中的诸多方面发挥着重要的作用,如智能电网、智能家居、智慧医疗、多媒体服务、气象预测与灾害预防、车联网等。由于智能终端设备的感知能力不断增强,其感知的数据规模以及数据处理的计算复杂性呈现出爆炸式的增长趋势,这导致传统的基于“中心交付方式”的云计算服务模式难以适应该趋势。据思科公司的数据预测,终端设备规模及其产生的数据规模均呈现出倍数级增长趋势,到2021年,全球范围内将有超过500亿的终端设备,这些设备每年产生的数据量将达到847 ZB。其中,约10%的数据需要通过计算处理,数据中心计算性能正逐渐达到瓶颈。相比而言,全球数据中心的存储能力预计仅能达到2.6 ZB,而网络流量为19.5 ZB。云计算模式下,所有的数据必须上传至集中式服务器,并在计算后返回相应的设备,终端设备数据的爆炸式增长也加剧了网络负载,严重影响服务质量,导致低响应时延、网络拥塞等问题。综上,集中式的云计算模式面临“算不动、存不下、传不畅”的挑战,难以满足终端环境爆炸式增长的数据处理需求。

为缓解数据中心的处理压力、消除计算与通信瓶颈、提升系统的服务质量,一种行之有效的方案是将云服务卸载到距离终端用户更近的位置,这就形成了一种新型的边缘计算范型,其基本架构如图1所示。实际上,终端设备的广泛部署在提升数据感知能力的同时,也为终端环境聚集了大量可用资源,其边缘计算及存储能力呈现倍数级增长。同时,万物互联时代的到来与边缘

网络中数据量的飞速增长,促进了终端设备间通信技术的进一步发展,涌现出新型的高数据率、低时延的通信模式,提高了边缘网络的网络传输容量。边缘设备通过通信网络基础设施实现设备互联,并构成一个泛在的边缘网络环境,实现数据的相互收集和交换。因此,利用边缘环境中终端设备的自有资源可以有效地缓解云计算中心的负载。

业界基于“将计算交付到数据感知源、就近提供智能服务”的思想,从不同的角度对边缘数据处理模型展开了广泛深入的研究。在学术界,针对物联网和内容分发网络(content distribution network, CDN)应用的高通信时延和高带宽成本问题,加州大学伯克利分校提出了微数据中心(micro data center)概念<sup>[1]</sup>,将小型化的数据中心部署得更靠近应用侧;针对移动计算环境中的网络时延和带宽不足的问题,卡内基梅隆大学提出了位于互联网边缘的微云(cloudlet)<sup>[2]</sup>,将其作为资源门户,为移动计算应用提供实时交互和云服务网关功能。在工业界,5G标准中正式将移动边缘计算<sup>[3]</sup>纳入其中,在接近移动用户的无线接入网范围内,提供信息技术服务和云计算能力;思科公司提出了雾计算(fog computing)<sup>[4]</sup>的概念,在云与移动设备之间引入“雾层”,扩展基于云的网络结构。美国国防高级研究计划局(Defense Advanced Research Projects Agency, DARPA)为应对战术环境中的信息实时处理和共享问题,提出了基于网络计算节点的分散计算(dispersed computing),它支持按需从周边节点借用计算资源和网络资源,从而实现任务的快速处理。

实际上,边缘计算核心理念可追溯到对等网络(peer to peer, P2P)计算和内容分发网络。作为一种计算泛型,边缘计算的基本思想是对边缘设备自有的分布式资源进行充分利用,即在数据感知源建立资源供

给机制,为用户提供就近服务,以缓解服务器/数据中心的压力,降低网络传输带宽消耗,并加快数据处理效率,增强服务响应能力。

在万物互联及大数据的环境下,虽然边缘计算范型为计算任务处理带来了新的机遇,但同时也带来了许多技术挑战。边缘环境中的计算任务处理通常需要边缘设备基于分布式协作完成,其核心是对边缘环境中计算、存储、网络资源的高效分配,以实现任务及数据的动态部署,从而在减少资源开销的前提下,满足用户对服务质量的需求。数据规模的增长对资源的协同优化提出了更高的要求。与此同时,边缘网络基础设施的弱连接特征(infrastructure-less)使得数据和任务的迁移面临更加复杂的网络环境,这加剧了分布式协同的复杂度。

考虑到边缘环境中的设备异构性、系统动态性,为了实现资源的高效整合,并适应用户多样化的服务需求,国内外研究者从计算、存储、网络等资源协同优化的角度展开研究,并产生了一系列开创性的进展。为了更好地理解边缘计算中相关的资源优化技术,本文首先介绍边缘计算范型和资源优化的基本理念,进一步,针对边缘环境中计算、存储、通信3种关键资源,从计算卸载、分布式缓存、高性能通信3个方面对现有工作进行总结,最后对未来的研究方向进行展望。

## 2 边缘计算基本范型与优势

### 2.1 边缘计算范型

基于云计算的数据处理框架(如MapReduce编程模型<sup>[5]</sup>、Hadoop分布式文件系统<sup>[6]</sup>、Spark内存计算框架<sup>[7]</sup>等)虽日趋成熟并被广泛应用,但这种集中式的数据处理架构难以适应爆炸性增长的边缘数据处理

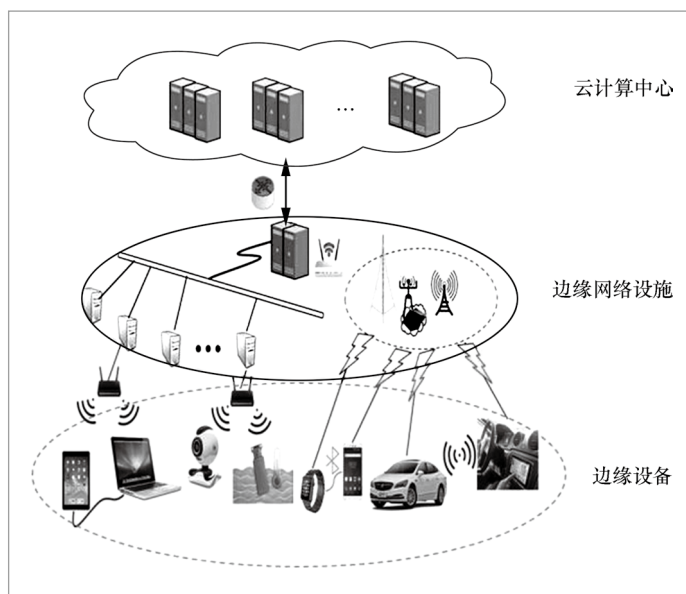


图1 边缘计算基本架构

需求。为了实现分散任务的高效处理,一种解决方案是利用云计算中心之间的协同。针对多地域分布云存储系统协同问题,通过设计基于文件相似性的协同传输机制,实现多平台间安全可靠的文件同步,并提高用户服务体验,如CoCloud系统<sup>[8]</sup>和Meta-Sync系统<sup>[9]</sup>。Chen L等人<sup>[10]</sup>研究了面向地域分布的多数据中心中的多任务调度技术,基于数据中心间带宽较低、稳定性差、数据在跨域数据中心间分布不均匀等特点,设计了一种保证公平性的作业调度算法。对于分布式存储这类对时延及计算要求较低的任务来说,多个云平台的协同对服务质量有一定的改善,但仍然无法解决边缘环境中广泛存在的时延敏感及计算密集型任务的服务质量保证问题。

为了保证海量边缘数据处理的实时性和有效性,业界先后提出了微数据中心、微云、雾计算、移动边缘计算等边缘数据处理模型,图2给出了这些边缘数据处理模型的基本架构。

#### (1) 微数据中心

微数据中心是一种处在网络边缘的小

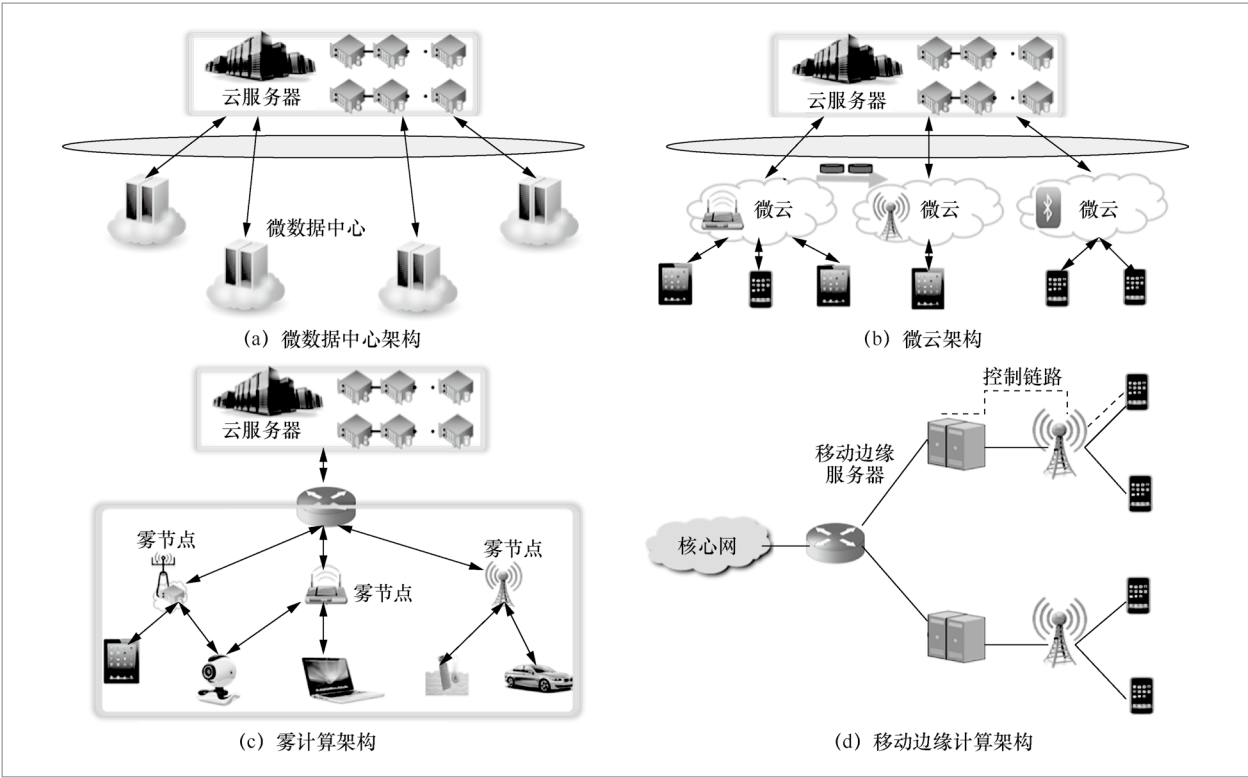


图 2 边缘数据处理模型的基本架构

型化或模块化的数据中心架构,其部署的位置更加靠近数据源,旨在降低面向物联网和内容分发网络的通信时延和带宽成本,同时解决不同类型的工作负载对传统云计算架构软硬件资源差异性需求的问题。由荷兰政府资助的、IBM公司和荷兰射电天文学研究所合作的DOME项目成功将微数据中心应用于全球规模最大的射电望远镜阵列项目,其中采用的每一个2U 19机架单元都整合了64台高性能服务器、网络、存储模块。Aazam M等人<sup>[11]</sup>将微数据中心引入物联网和云平台之间,旨在对物联网多类型服务提供资源管理、数据过滤、预处理以及安全措施。

(2) 微云

微云处在云平台和终端设备之间的中间层次,是被部署在网络边缘、具有移动性的小型数据中心。微云旨在支持资源密集型

和实时交互的移动应用,如增强现实应用、远程渲染的视频游戏等。通常微云部署的位置与终端用户的距离为一跳无线连接,比如部署在蜂窝网基站或者Wi-Fi基站上,为终端用户的计算任务提供低时延响应。多个微云构成分布式的移动边缘计算环境,拓展用户可用资源,通过提供类似云平台的动态迁移机制,实现资源的负载均衡。Chen M等人<sup>[12]</sup>提出一种由边缘移动设备基于短距离无线链路互联而成的微云架构,每一个移动设备都可以作为计算服务的提供者或者请求者。在这种由多个微云自组织连接构成的网络中,计算任务可以选择本地执行、直连的微云执行、基于机会路由传输给其他微云执行。Jia M等人<sup>[13]</sup>为了满足跨不同计算平台构建无缝衔接的应用程序体验需求,提出了基于边缘网络中多个微云工作负载均衡的优化,降低了处理结果的返回时延。



### (3) 雾计算

思科公司在2012年提出雾计算概念,并在2014年发布供研发者使用的雾计算开发套件IOx。随后,思科公司联合ARM公司、戴尔公司、英特尔公司、微软公司和普林斯顿大学,于2015年联合成立了开放雾计算联盟(OpenFog Consortium)。雾计算节点由性能相对较弱但地理位置广泛分散的雾节点构成。作为云计算的延伸,雾计算强调终端用户与计算承载节点的接近度、时延和骨干网带宽开销、自适应移动性和位置感知等,以实现更好的服务质量体验和边缘数据特征分析,同时缓解云处理中心的计算负载。近年来,国内外研究者从基本架构、资源管理、任务调度、节能、数据安全等多个方面对雾计算展开了研究,并取得了富有前景的研究成果<sup>[14-15]</sup>。

### (4) 移动边缘计算

移动边缘计算最早由欧洲电信标准协会(European Telecommunications Standards Institute, ETSI)提出,其核心思想是在无线电接入网(radio access network, RAN)的基站中提供计算及存储资源,使得能在更近的位置响应用户的计算任务和数据请求<sup>[3]</sup>。此外,移动边缘计算带动了一系列相关领域研究的飞速发展,并正式纳入5G标准<sup>[16]</sup>。移动边缘计算方向也吸引了更多的学者进行大量深入的研究,并与移动智能设备的万物互联网理念相结合,产生了众多富有前景的技术飞跃和广泛部署的应用平台<sup>[17-18]</sup>。

## 2.2 边缘环境资源整合优势

边缘环境中包含大量的多样性边缘设备,这些边缘设备通常具有一定的计算能力和存储能力,同时提供了多个通信接口。以移动边缘设备为例,它们可以通过蜂窝网链路直接连接到宏蜂窝基站,同时基于Wi-

Fi、蓝牙等通信模式构建设备到设备的自组织通信网络。这些设备空闲的计算、存储及通信资源在移动边缘网络中构成一个巨大的资源池,通过缓存一些计算功能和数据,大量的计算任务和数据请求可以被卸载到由移动边缘设备构成的自组织网络中。通过协同利用这些边缘资源,可以实现边缘设备任务的快速响应以及数据的高效分发。另外,边缘环境中包含大量分布式的微型数据中心、部署在宏蜂窝基站上的移动边缘服务器以及家庭基站和微型基站等边缘网络设备,可以有效地缓解集中式数据中心的计算负载和流量负载。此外,密集部署的边缘网络设备与宏基站协同形成了无线网络的全面覆盖,有效地提升了边缘网络的容量。这种多样性密集部署的无线网络正是5G中网络致密化(network densification)的思想,是研究未来网络架构的一个重要趋势。

边缘计算有机地融合了分散在网络终端上的计算、存储及通信资源,与传统宏基站集中式网络架构相比,移动边缘网络主要具有以下3点优势。

- 低时延: 由于数据存储和数据处理位于距离终端用户更近的位置,通信的传输时延可以被有效地降低。其中最典型的受益案例是计算任务卸载和视频边缘缓存。在计算任务卸载场景中,边缘云和边缘服务器通常可以为终端用户提供更高质量的数据传输,同时避免集中式服务造成网络拥塞。在视频服务的边缘缓存中,将视频部署在移动边缘设备,响应流媒体用户的服务请求,可以有效地避免视频在核心网络和宏基站回程链路(backhaul link)的时延,提高视频的服务质量。

- 低能量开销: 之前的研究工作表明,与集中式任务处理相比,将计算任务卸载到雾计算节点和微云上,可以有效地降低能量开销<sup>[19-20]</sup>。此外,针对流媒体这类数据请求

型服务, 在边缘网络设备或移动智能设备上缓存数据并响应请求, 可以有效地避免数据在核心网或基站回程链路上的冗余传输, 而边缘智能设备或网络终端缓存单位比特的数据所需的能量开销要远低于传输带来的能量开销。因此, 对计算、存储、通信资源的合理分配利用, 可以在整个网络层面实现绿色节能计算。

● 内容感知: 在边缘环境中, 计算任务以及数据请求模式通常具有一定的地域规律性。边缘网络中的大量的终端设备距离移动用户更近, 同时具有空余的资源, 能够方便高效地统计、学习出特定区域内用户的计算任务请求模式、移动模式、行为模式以及视频数据的请求概率等。这些信息对网络中的流量监控、路由策略、资源分配调度等具有很好的指导意义, 据此网络提供商和数据中心可以提高自身的资源利用效率, 同时保证对终端用户的服务质量。

3 边缘计算资源优化技术

3.1 面向计算资源优化的计算卸载

基于云计算的集中式数据处理技术已经不能及时高效地处理边缘设备产生的数据, 根本原因在于线性增长的集中式云计算能力无法匹配倍数级增长的边缘数据, 造成数据处理任务的排队等待。与此同时, 云数据中心有限的存储能力和带宽能力使得集中式处理计算任务的性能进一步恶化。如图3所示, 基于边缘计算模型的理念, 可以将计算卸载到距离用户更近的位置, 以解决面向万物互联时代的边缘端大数据处理效率问题。通过建立广泛分布的移动边缘云、微云、微数据中心等, 任何位置的用户都可以方便快捷地使用边缘云服务, 有效地缓解数据中心的计算负载和带宽开销, 同时提升服务的响应能力。在工业界, 各个云服务提供商在部署集中式数据中心的同时, 也加快了分布式的边缘云设施的建设, 如阿里云的内容分发网络在全球部署了超过1 500个边缘服务器。

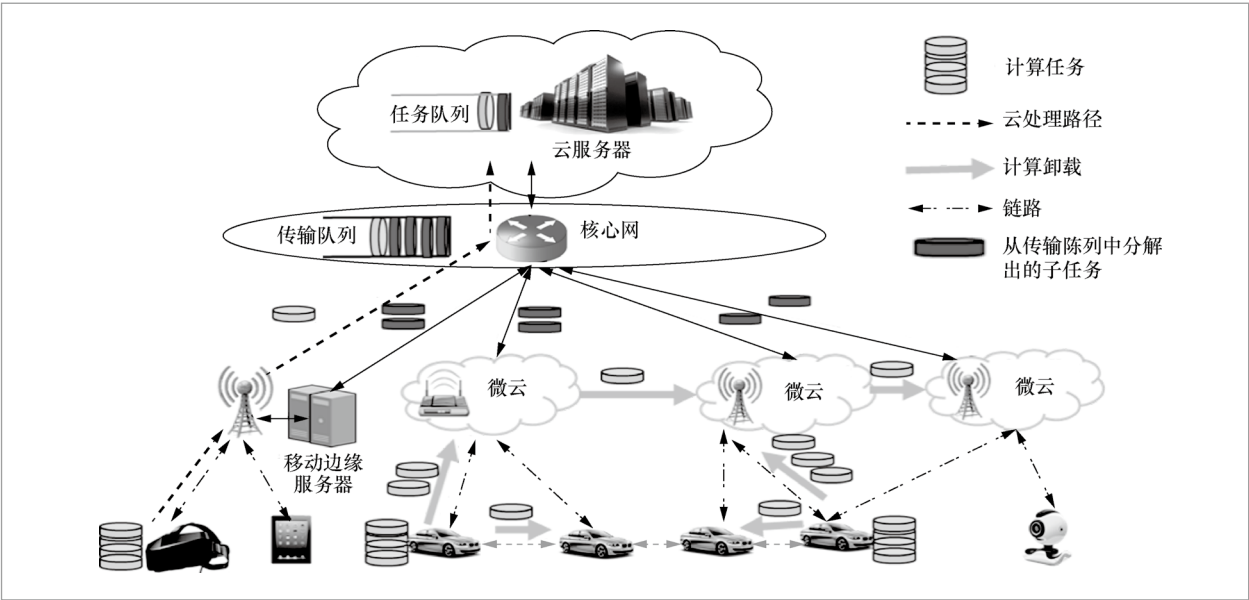


图3 边缘环境中的计算卸载模式

在考虑终端设备产生的计算卸载时, 主要面临两个关键问题: 哪些计算需要被卸载到边缘环境中执行; 将计算卸载到哪些边缘节点执行。对于第一个问题, 以移动设备上的视频游戏、增强现实这类计算为例进行说明。受限于自身的计算能力和电池能力, 移动设备难以保证持续良好的质量体验。而集中式云计算中心虽有强大的计算能力, 但由于可能存在的任务排队以及网络拥塞, 难以提供稳定的较低往返时延。一种行之有效的方案是, 将计算卸载到距离移动设备较近的边缘服务器。对于第二个问题, 当有多个计算任务需要卸载到边缘节点时, 盲目的卸载很容易导致部分边缘节点的负载不均衡。同时, 考虑到计算卸载时涉及的数据传输, 选择边缘节点时也需考虑与设备之间的网络状况。

Chen X等人<sup>[21]</sup>考虑了多信道无线干扰环境中面向多用户的计算卸载问题。由于多个用户对信道资源以及计算资源存在竞争关系, 作者通过博弈模型设计了分布式的计算卸载算法, 证明该算法可以在收敛的时间限制下得出纳什均衡, 并量化分析该算法与集中式解决方案之间的竞争比。同样基于多用户的计算卸载问题, Deng M等人<sup>[22]</sup>采用博弈的方法, 基于用户的通信和计算资源, 顺序执行计算任务的卸载。You C等人<sup>[23]</sup>研究了密集的移动计算向基于蜂窝网的边缘云卸载的问题。在基于时分多址 (time division multiple access, TDMA) 的用户接入以及计算等待时间和移动设备能量开销的约束下, 作者将卸载问题刻画为一个凸优化问题, 并给出最优解; 在基于正交频分多址 (orthogonal frequency division multiple access, OFDMA) 的接入方式下, 将其刻画为混合整数规划问题, 并给出近似最优解。Mao Y等人<sup>[24]</sup>研究将密集型计算卸载到移动边缘服务器的问题。为了实现绿色节能计算, 作者采用带有能量收集系统的移动

边缘服务器设备, 通过决定是否卸载、边缘服务器CPU执行频率以及无线传输的发射功率, 尽可能地降低计算的执行等待时间。基于移动边缘计算, Wang Y等人<sup>[25]</sup>在计算卸载中引入动态电压缩放 (dynamic voltage scaling, DVS) 技术。基于计算负载以及性能需求, DVS技术支持设备上的电压和时钟频率动态调整, 可以实现更加弹性的计算卸载策略。Wang F等人<sup>[26]</sup>考虑支持无线供电的多用户移动边缘计算场景, 多个无线接入点向端用户提供无线供电, 端用户基于此执行计算任务或将计算卸载到无线接入点上。作者提出了一种最佳资源分配方案, 在计算时延约束下, 最小化无线接入点的能量开销。

另外, 为了适应数据规模的增长速度, 数据处理方式正由传统的在单个机器上进行集中处理逐渐过渡到在由众多小规模不可靠机器构成的大规模分布式系统上处理。例如在MapReduce框架中, 需要将计算任务分散到多个节点上执行, 同时节点之间需要大量的数据交换。传统的集群系统中存在的节点失效、计算落后、通信瓶颈问题影响着整体的计算性能。这些问题可以通过编码技术解决, 即通过编码方法生成冗余计算任务, 并保证最终计算结果能够从任意足够数目的节点上的计算结果中获得<sup>[27-28]</sup>。节点计算、网络资源的异构性在边缘计算环境中更加显著, 研究者也将这一类编码计算方法推广到边缘计算的框架下, 以实现计算资源的高效利用以及与存储、通信资源的统一优化。Aliasgari M等人<sup>[29]</sup>将编码方法运用到C-RAN网络架构中, 以保证云端基于网络功能虚拟化的上行信号的快速解码。Li S等人<sup>[30]</sup>对编码方法在边缘计算的应用进行了前瞻性分析, 并提出了一种通用的编码边缘计算方法, 其中边缘计算节点能够利用计算的编码结果生成干扰信号迫零的通信消息, 从而提升无线频谱利用率<sup>[31]</sup>。

3.2 面向存储资源优化的边缘缓存

云数据中心通常采用大规模存储集群对数据进行集中式的存储管理, 存储能力大, 并且存储节点之间的网络性能较好。同时, 分布式存储系统通常通过数据冗余存储应对失效节点, 保证数据的可用性。学术界提出, 在存储系统中采用基于编码的分布式容错机制, 能够有效地降低存储开销, 并提升系统的整体容错能力, 对冗余编码数据的解码机制能够保证在部分节点失效时可以进行数据恢复<sup>[32-34]</sup>。然而, 只依靠数据中心的存储模式仍然无法满足倍数级增长的边缘数据的存储需求。一种方案是在云存储系统中引入边缘节点的协助。Zhao J等人<sup>[35]</sup>提出一种基于反向拍卖机制的边缘节点协助存储模型, 边缘节点将自有的存储资源当作商品拍卖给云存储平台, 通过诚信拍卖机制保证整体收益的最大化, 提高存储资源的利用效率。Zhao W等人<sup>[36]</sup>考虑云存储系统中的流媒体服务平台, 引入边缘节点/雾节点, 通过边缘节点贡献的存储资源和带宽资源, 并辅以最优副本分布的计算, 提高边缘存储资源的利用效率。

随着边缘数据规模的不断增长, 传统的基于云数据中心的存储架构会产生数据

中心负载过重、网络拥塞等问题。云平台提供商将面临巨额的带宽开销及存储开销, 同时边缘用户的服务质量也难以保证。为了解决这些问题, 一种行之有效的方案是边缘的分布式缓存技术, 通过在边缘网络上的缓存设备, 将热门的流媒体数据预先缓存到距离用户更“近”的网络设备上。由于边缘环境中设备的多样性和异构型, 数据缓存在网络中的不同位置将产生不同的开销与收益。图4展示了在边缘网络的不同位置部署缓存。例如, 在CDN服务器上的数据缓存可以考虑数据请求的地域特征, 通过在CDN服务器部署数据缓存以及服务器内部的数据迁移实现数据请求的就近响应, 降低终端用户的请求开销<sup>[37]</sup>。而在面向移动终端的网络环境中, 考虑到边缘设备存储能力的限制, 通常采用主动式的缓存方式将热门的数据进行预先部署, 同时将数据分别缓存在不同的层次位置, 如无线接入网层、边缘移动设备层等, 在不同层次上避免传输冗余<sup>[38]</sup>。Abani N等人<sup>[39]</sup>将边缘缓存扩展到网络中的各个层面, 通过预测用户移动性并结合信息中心网络, 确定用户数据请求在网络中的最佳预取节点, 降低响应时延。

在RAN层, Ahlehigh H等人<sup>[40]</sup>提出基于无线接入网基站的视频分布式缓存机

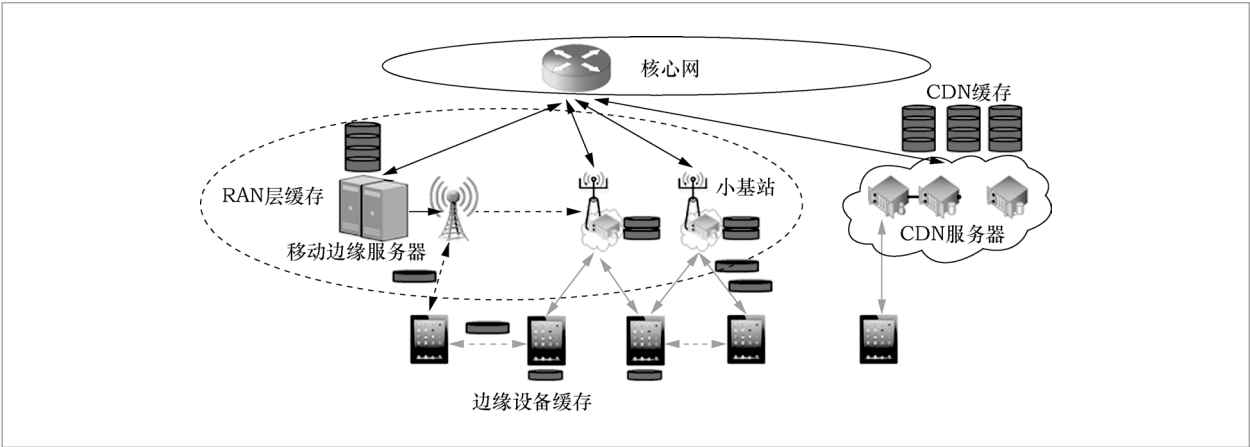


图4 边缘网络中分布式缓存的各个层次



制,显著提高了边缘网络环境能够承载的视频容量以及用户的视频播放体验。由于无线接入网基站与CDN服务器相比,具有较小的缓存空间,作者提出了基于活跃用户感知的主动式缓存策略,同时基于回程链路感知和无线信道的调度技术,设计边缘缓存策略,以确保边缘网络设备同时支持的视频的传输个数最大化。Gomes A S等人<sup>[41]</sup>考虑在移动边缘计算环境中融合下一代网络架构以及信息中心网络 (information centric networking, ICN),通过预测用户的移动模式,设计了高效率的缓存迁移算法,提高了用户数据请求在边缘的命中效率。Tandon R等人<sup>[42]</sup>在新型的雾-无线接入网 (fog-RAN) 架构下,通过引入标准化交付时间这一性能评价指标,确定边缘缓存的最佳放置以及协同传输策略。

另外,小型蜂窝网 (如家庭基站和微型基站) 的网络密集化是未来5G网络架构的一个重要思路,这些小基站具有传输代价低、时延低、覆盖范围广泛等特点,也是边缘环境中应对爆炸式数据需求的富有前景的方法。Zhang J等人<sup>[43]</sup>在密集的异构无线网络中引入软件定义网络 (software defined network, SDN) 技术,同时通过边缘缓存技术,实现比当前LTE网络更高的吞吐量和改进的能耗。另一种缓存策略的设计思路是,基于数据流行度以及用户行为的预测、统计,设计数据在小基站上的缓存方案,以提高用户请求在小基站上的命中率,减轻移动蜂窝网络中回程链路的通信压力,提升系统吞吐量。Golrezaei N等人<sup>[44]</sup>首次提出在家庭基站部署缓存的思想,并设计了最小化请求时延的缓存部署优化算法。Poularakis K等人<sup>[45]</sup>进一步考虑了基站的通信带宽限制,并针对用户移动设备的通信能耗指标,提出了一种具有性能保证的低复杂度的启发式算法。Gharaibeh A等人<sup>[46]</sup>考虑了小型基站之间的协同缓存机制,并设

计了一种高效的在线合作式缓存部署与替换方法。Avrachenkov K等人<sup>[47]</sup>将小基站上的缓存部署问题等价于博弈论中的潜在博弈 (potential game) 问题,通过寻找潜在博弈的纳什均衡找到全局的最优缓存部署策略。Li J等人<sup>[48]</sup>在工作中考虑网络设备提供商与视频服务提供商之间的竞争关系,通过双寡头博弈 (Stackelberg game) 模型刻画缓存设备租用及部署策略,从而同时最大化网络设备提供商和视频服务提供商的收益。Ma G等人<sup>[49]</sup>则充分考虑了可能影响到缓存效能的一些关键因素,通过真实的数据,说明视频的热门度分布、用户的移动行为及请求模式、小基站缓存策略对服务质量的影响。Dehghan M等人<sup>[50]</sup>考虑了内容在互联网和蜂窝网的跨层缓存与分发机制,基于后端服务器路径拥塞敏感和不敏感两种情形,联合优化了内容的缓存部署问题和用户请求的路由选择问题。Shnaiwer Y等人<sup>[51]</sup>针对文件下载场景,基于机会网络编码 (opportunistic network coding) 技术,卸载蜂窝宏基站的流量到家庭基站网络上,同时提高移动用户的质量体验。在这项研究工作中,家庭基站上文件的部署问题被转化为机会网络编码图优化问题。Hajri S E等人<sup>[52]</sup>则考虑小基站网络上主动式缓存部署的能耗问题,通过对移动用户空间位置属性以及视频请求概率的聚类分析,设计了小基站上视频的缓存部署策略,以提高用户请求的命中率和整个系统的能量效率。

在边缘移动设备中,通过利用移动设备自有的存储资源,同时结合设备到设备 (device to device, D2D) 的通信模式,业界提出在用户移动设备之间建立协作缓存机制,可有效提升系统吞吐量,降低集中式数据中心的通信开销,同时减少边缘移动设备的蜂窝网流量开销。Golrezaei N等人<sup>[53]</sup>提出一种基于D2D通信的边缘设备缓存架构,边缘设备可以通过功率控制改变设备

间的协作距离。作者指出了设备之间的协作距离与设备间传输干扰之间的矛盾, 并通过功率控制算法实现了频谱利用最大化, 有效提升了系统吞吐量。Ji M等人<sup>[54]</sup>面向基于传统的微波(2 GHz)和毫米波(mmWave)构建的D2D网络, 设计了基于边缘设备的缓存系统, 并且验证了在异步数据需求情形下缓存可以提高系统吞吐率。Yang C等人<sup>[55]</sup>在异构无线网络中, 基于协作的用户缓存对系统吞吐量及用户访问时延进行了理论分析。Maddah-Ali M A等人<sup>[56]</sup>则发现将无线基站的广播特性与网络编码机制结合起来, 可以显著提升缓存效率。此外, 由于基于移动边缘设备的缓存机制需要用户贡献自有资源, Chaintreau A等人<sup>[57]</sup>探讨了基于用户博弈的激励机制, 通过社会福利最大化, 给出了移动边缘网络中的分布式缓存策略。

3.3 面向通信资源优化的高性能协同传输

如图5所示, 边缘计算环境中包含多样性的通信方式, 与传统的云处理架构相比,

基于毫米波的通信技术和基于移动边缘设备之间的D2D通信技术是较具特色的两类传输模式。毫米波代表了5G网络中先进的通信技术, 极大地丰富了网络带宽。D2D通信基于设备之间的直连链路, 极大地提高了信道资源的利用效率, 提升了边缘网络的容量。此外, 边缘环境中的数据传输通常与边缘设备的缓存密切相关, 在很多情况下, 传输优化会与缓存优化相结合。基于此, 本节从毫米波通信、D2D通信、基于缓存的协同通信技术3个方面对相关工作进行总结。

(1) 毫米波通信

毫米波通信具有极宽的带宽和极高的传输质量, 特别适用于5G架构下针对大规模数据的传输任务。由于毫米波的天线方向性和终端用户的高度移动性, 用户在数据传输过程中经常会遭受较长的切换时延和连接时延。为了保证毫米波定向天线覆盖区域内视频用户的服务质量, Qiao J等人<sup>[58]</sup>将视频传输问题转化为高速缓存的分配问题, 并基于马尔可夫决策过程为每个用户动态地分配缓

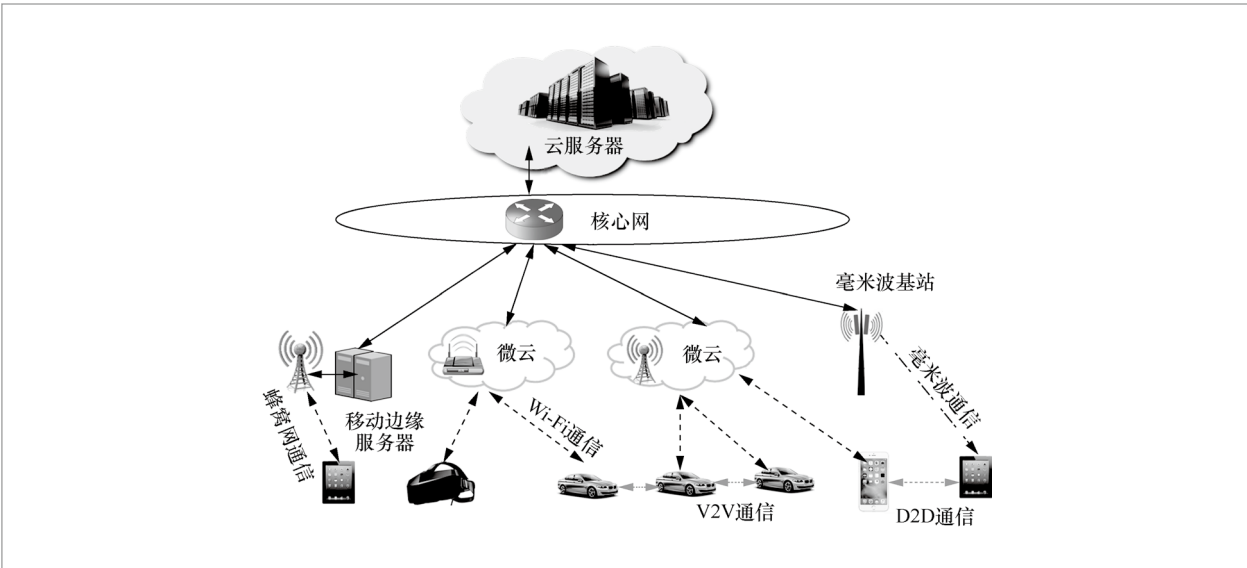


图 5 边缘设备支持的多样性通信模式

存。此外, 严苛的传播条件限制了毫米波在蜂窝网络中的接入性能。由于定向传输必须基于蜂窝发现和同步的过程, 在多个方向上的多个传输会产生一定量的时延。Filippini I 等人<sup>[59]</sup>提出了一种基于用户位置感知的蜂窝发现策略, 同时随时记录障碍信息, 为后续的蜂窝发现提供指导。

## (2) D2D通信

位置接近的移动边缘设备可以通过网络接口构建直连的D2D通信链路。由于这种通信方式具有高频谱利用率、低能量开销、信道资源复用等特点, 近年来得到了广泛的关注。Kella L 等人<sup>[60]</sup>实现了基于单跳D2D链路的视频文件传输平台, 基于用户之间的合作原则, 利用网络编码技术和边缘设备的广播性能, 实现了网络效用最大化。在移动边缘云场景中, Abdelwahab S 等人<sup>[61]</sup>设计了协同传输模型, 以支持服务和应用向移动边缘设备卸载。通过对无线网络环境的感知以及利用D2D通信技术, 作者提出了Replisom架构, 基于压缩采样理论中的稀疏恢复技术, 解决多个边缘设备同时传输时对无线信道资源的竞争问题, 有效地降低了任务卸载过程中的时延和能量开销。针对移动边缘环境中无处不在的数据传输以及时延敏感型任务, Nunna S 等人<sup>[62]</sup>结合5G和移动边缘计算架构, 提出了一个具有超低时延且基于内容感知的传输模型。Cao W 等人<sup>[63]</sup>在将蜂窝网流量卸载到D2D网络中时, 考虑到基站与移动边缘设备之间的位置关系, 提出了一种基于整数规划的流量卸载方案, 并且证明了这种传输方案在能耗上的优越性。

## (3) 基于缓存的协同通信技术

结合边缘缓存的协同传输模式可以有效地提升数据的传输效率。然而, 考虑到网络的高度动态性、用户规模的急剧增长以及网络传输资源的有限性等因素, 设计与边缘网络缓存相匹配的高效传输技术面临着更

大的挑战。从结合边缘网络设备缓存的角度, Ao W C 等人<sup>[64-65]</sup>结合无线网络数据传输模型, 为热门数据在小基站上部署了多个副本, 以保证协同传输时达到最高的数据率, 同时在小基站上部署冷门数据, 保证所有媒体数据的多样性与可用性。Khreishah A 等人<sup>[66]</sup>则考虑小基站同时传输数据时的干扰, 结合信道分配算法, 设计相应的小基站缓存部署策略, 提高了整体的分发效率。从结合未来网络架构中缓存策略的角度来看, Yuan X 等人<sup>[67]</sup>和Kanai K 等人<sup>[68]</sup>则分别考虑在路由器上和交通工具上部署缓存设备, 通过复用内容中心网络 (content centric network, CCN) 及命名数据网络 (named data network, NDN) 的传输协议, 结合缓存部署策略, 高效地分发流媒体数据。此外, Zhao N 等人<sup>[69]</sup>对5G网络架构中边缘网络设备之间, 尤其是大量部署的小基站之间的干扰控制技术面临的挑战进行了总结, 并展望了利用缓存和计算资源简化网络拓扑、提高吞吐量的优化技术。Andreev S 等人<sup>[70]</sup>结合用户的移动模式和具体的应用需求, 揭示了5G网络中融合通信、计算、缓存的关键要素, 并面向网络运营商的收益给出相应的优化模型。

# 4 结束语

本文对面向边缘计算的资源优化技术的进展进行了总结。边缘计算的理念充分利用边缘端丰富的资源, 为终端用户提供便捷的资源接入与高效的计算任务响应, 卸载集中式云计算中心的负载。从边缘资源优化的角度, 本文分别总结了面向计算资源优化的计算卸载方法、面向存储资源优化的分布式缓存机制以及面向通信资源优化的高性能协同传输技术。虽然现有的工作对这一领域展开了广泛的研究, 并取得了一定的进展, 但是这些工作大多数关注于边缘资源的优化, 忽略

了对云平台资源的有机结合与协同。此外,多应用复杂环境下的资源调度以及面向真实环境的资源优化方法仍然是边缘计算中亟须解决的关键问题。

未来的研究需要在“云中心-网络-边缘端”3个层面的有机融合场景下实现更大范围的资源协同,同时建立面向云中心与边缘端的双向的资源协同优化理论和面向实际环境的资源调度方法。未来的研究方向主要总结为以下几个方面。

(1) 有机融合云计算中心与边缘计算环境,实现更高效的资源利用

云计算与边缘计算具有各自的性能优势,同时可以相互弥补对方的不足。叶保留等人<sup>[71]</sup>在中国计算机学会(CCF)《2017—2018中国计算机科学技术发展报告》中展望了基于云中心与边缘端的融合计算模型的未来研究趋势。将云计算与边缘计算进行有机融合是当前发展的必然趋势。然而,现有大多数资源优化方面的研究工作集中在对云资源、边缘资源中某一方面进行优化,难以有效应用于云端融合。如何针对不同的应用需求,兼顾云数据中心与边缘环境各自的特点,优化资源的使用,仍面临巨大的技术挑战。

(2) 面向多样并发服务,构建高效、负载均衡的动态资源调度机制

现有大多数边缘计算资源调度方面的工作主要针对特定应用展开,围绕应用服务需求建立资源调度机制。然而,边缘计算环境需要支持多种多样的应用服务,而这些服务对资源的需求各异,且资源间存在共享或竞争的关系。此外,现有资源调度方法较少地考虑了边缘环境整体的负载均衡。如何面向多样并发服务,构建高效、负载均衡的动态资源调度机制是边缘计算中亟须解决的一个关键问题。鉴于边缘环境的复杂性与服务的多样性,基于机器学习的调度方法值得尝试。

(3) 面向真实环境,建立实际可用、可靠高效的资源调度方法

目前大多数边缘计算资源调度方面的基础理论研究对边缘环境进行了或多或少的假设,而这些假设仍有待进行实证研究与验证。笔者认为,在边缘计算研究背景下,需要进一步从系统层面出发,指导资源调度理论研究工作的推进,在此基础上进行大量的实验验证和经验总结,弥补基于理想环境的资源调度理论与真实环境调度机制设计之间的鸿沟,从而建立实际可用、可靠高效的资源调度方法。

## 参考文献:

- [1] ARMBRUST M, FOX A, GRIFFITH R, et al. A view of cloud computing[J]. Communications of the ACM, 2010, 53(4): 50-58.
- [2] GAI K, QIU M, ZHAO H, et al. Dynamic energy-aware cloudlet-based mobile cloud computing model for green computing[J]. Journal of Network and Computer Applications, 2016, 59(C): 46-54.
- [3] HU Y C, PATEL M, SABELLA D, et al. Mobile edge computing-a key technology towards 5G[R]. [S.l.]: ETSI, 2015.
- [4] BONOMI F, MILITO R, ZHU J, et al. Fog computing and its role in the Internet of things[C]//The 1st Edition of the MCC Workshop on Mobile Cloud Computing, August 17, 2012, Helsinki, Finland. New York: ACM Press, 2012: 13-16.
- [5] DEAN J, GHEMAWAT S. MapReduce: simplified data processing on large clusters[C]//The 6th Conference on Symposium on Operating Systems Design and Implementation, December 6-8, 2004, San Francisco, USA. Berkeley: USENIX Association, 2004: 137-150.
- [6] SHVACHKO K, KUANG H, RADIA S, et al. The hadoop distributed file system[C]//2010 IEEE 26th Symposium on Mass Storage Systems and Technologies, May 3-7, 2010, Incline Village, USA. Piscataway: IEEE Press, 2010:



- 1-10.
- [7] ZAHARIA M, CHOWDHURY M, FRANKLIN M J, et al. Spark: cluster computing with working sets[C]// The 2nd USENIX Conference on Hot Topics in Cloud Computing, June 22-25, 2010, Boston, USA. Berkeley: USENIX Association, 2010.
  - [8] E J L, CUI Y, WANG P, et al. CoCloud: enabling efficient cross-cloud file collaboration based on inefficient web APIs[C]//IEEE Conference on Computer Communications (INFOCOM), May 1-4, 2017, Atlanta, USA. Piscataway: IEEE Press, 2017: 1-9.
  - [9] HAN S, SHEN H, KIM T, et al. MetaSync: file synchronization across multiple untrusted storage services[C]//The 2015 USENIX Annual Technical Conference (ATC), July 8-10, 2015, Santa Clara, USA. Berkeley: USENIX Association, 2015: 83-95.
  - [10] CHEN L, LIU S, LI B, et al. Scheduling jobs across geo-distributed datacenters with max-min fairness[C]//IEEE Conference on Computer Communications (INFOCOM), May 1-4, 2017, Atlanta, USA. Piscataway: IEEE Press, 2017.
  - [11] AAZAM M, HUH E N. Fog computing micro datacenter based dynamic resource estimation and pricing model for IoT[C]//IEEE 29th International Conference on Advanced Information Networking and Applications, March 24-27, 2015, Gwangju, Korea. Piscataway: IEEE Press, 2015: 687-694.
  - [12] CHEN M, HAO Y, LI Y, et al. On the computation offloading at ad hoc cloudlet: architecture and service modes[J]. IEEE Communications Magazine, 2015, 53(6): 18-24.
  - [13] JIA M, LIANG W, XU Z, et al. QoS-Aware cloudlet load balancing in wireless metropolitan area networks[J]. IEEE Transactions on Cloud Computing, 2018(1): 1.
  - [14] MAHMUD R, KOTAGIRI R, BUYYA R. Fog computing: a taxonomy, survey and future directions[M]//In Internet of everything. Heidelberg: Springer, 2018: 103-130.
  - [15] MOURADIAN C, NABOULSI D, YANGUI S, et al. A comprehensive survey on fog computing: state-of-the-art and research challenges[J]. IEEE Communications Surveys and Tutorials, 2018, 20(1): 416-464.
  - [16] 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.
  - [17] ABBAS N, ZHANG Y, TAHERKORDI A, et al. Mobile edge computing: a survey[J]. IEEE Internet of Thing Journal, 2018, 5(1): 450-465.
  - [18] 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: 6757-6779.
  - [19] GAO Y, HU W, HA K, et al. Are cloudlets necessary[R]. [S.l.: s.n.], 2015.
  - [20] JALALI F, HINTON K, AYRE R, et al. Fog computing may help to save energy in cloud computing[J]. IEEE Journal on Selected Areas in Communications, 2016, 34(5): 1728-1739.
  - [21] CHEN X, JIAO L, LI W, et al. Efficient multi-user computation offloading for mobile-edge cloud computing[J]. IEEE/ACM Transactions on Networking, 2016, 1(5): 2795-808.
  - [22] DENG M, TIAN H, LYU X. Adaptive sequential offloading game for multi-cell mobile edge computing[C]//The 23rd IEEE International Conference on Telecommunications (ICT), May 16-18, 2016, Thessaloniki, Greece. Piscataway: IEEE Press, 2016.
  - [23] YOU C, HUANG K, CHAE H, et al. Energy-efficient resource allocation for mobile-edge computation offloading[J]. IEEE Transactions on Wireless Communications, 2017, 16(3): 1397-1411.
  - [24] MAO Y, ZHANG J, LETAIEF B K. Dynamic computation offloading for mobile-edge computing with energy harvesting devices[J]. IEEE Journal on Selected Areas in Communications, 2016, 34(12): 3590-3605.
  - [25] WANG Y, SHENG M, WANG X, et al. Mobile-edge computing: partial computation offloading using dynamic voltage scaling[J]. IEEE Transactions on Communications, 2016, 64(10): 4268-4282.
  - [26] WANG F, XU J, WANG X, et al. Joint

- offloading and computing optimization in wireless powered mobile-edge computing systems[J]. IEEE Transactions on Wireless Communications, 2018, 17(3): 1784-1797.
- [27] LEE K, LAM M, PEDARSANI R, et al. Speeding up distributed machine learning using codes[J]. IEEE Transactions on Information Theory, 2018, 64(3): 1514-1529.
- [28] LI S, MADDAAH-ALI M A, YU Q, et al. A fundamental tradeoff between computation and communication in distributed computing[J]. IEEE Transactions on Information Theory, 2018, 64(1): 109-128.
- [29] ALIASGARI M, KLIEWER J, SIMEONE O. Coded computation against straggling decoders for network function virtualization[C]//IEEE International Symposium on Information Theory (ISIT), June 17-22, 2018, Vail, USA. Piscataway: IEEE Press, 2018.
- [30] LI S, MADDAAH-ALI M A, AVESTIMEHR A S. Coding for distributed fog computing[J]. IEEE Communications Magazine, 2017, 55(4): 34-40.
- [31] LI S, MADDAAH-ALI M A, AVESTIMEHR A S. Communication-aware computing for edge processing[C]//IEEE International Symposium on Information Theory (ISIT), June 25-30, 2017, Aachen, Germany. Piscataway: IEEE Press, 2017: 2885-2889.
- [32] DIMAKIS A G, GODFREY P B, WU Y, et al. Network coding for distributed storage systems[J]. IEEE Transactions on Information Theory, 2010, 56(9): 4539-4551.
- [33] SILBERSTEIN N, RAWAT A S, VISHWANATH S. Error resilience in distributed storage via rank-metric codes[C]//The 50th Annual Allerton Conference on Communication, Control, and Computing, October 1-5, 2012, Monticello, USA. Piscataway: IEEE Press, 2012: 1150-1157.
- [34] RAWAT A S, KOYLUOGLU O O, SILBERSTEIN N, et al. Optimal locally repairable and secure codes for distributed storage systems[J]. IEEE Transactions on Information Theory, 2014, 60(1): 212-236.
- [35] ZHAO J, CHU X, LIU H, et al. Online procurement auctions for resource pooling in client-assisted cloud storage systems[C]//2015 IEEE Conference on Computer Communication (INFOCOM), April 26-May 1, 2015, Hong Kong, China. Piscataway: IEEE Press, 2015.
- [36] ZHAO W, LIU J, HARA T. Optimal replica distribution in edge-node-assisted cloud-P2P platforms for real-time streaming[J]. IEEE Transactions on Vehicular Technology, 2018, 67(9): 8637-8646.
- [37] MUKERJEEK, NAYLORD, JIANG J, et al. Practical, real-time centralized control for CDN-based live video delivery[C]//The 2015 ACM Conference on Special Interest Group on Data Communication, August 17-21, 2015, London, UK. New York: ACM Press, 2015: 311-324.
- [38] WANG X, CHEN M, TALEB T, et al. Cache in the air: exploiting content caching and delivery techniques for 5G systems[J]. IEEE Communication Magazine, 2014, 52(2): 131-139.
- [39] ABANI N, BRAUN T, GERLA M. Proactive caching with mobility prediction under uncertainty in information-centric networks[C]//The 4th ACM Conference on Information-Centric Networking, September 26-28, 2017, Berlin, Germany. New York: ACM Press, 2017: 88-97.
- [40] AHLEHAGH H, DEY S. Video-aware scheduling and caching in the radio access network[J]. IEEE/ACM Transactions on Networking, 2014, 22(5): 1444-1462.
- [41] GOMES A S, SOUSA B, PALMA D, et al. Edge caching with mobility prediction in virtualized LTE mobile networks[J]. Future Generation Computer Systems, 2017, 70: 148-162.
- [42] TANDON R, SIMEONE O. Cloud-aided wireless networks with edge caching: fundamental latency trade-offs in fog radio access networks[C]//IEEE International Symposium on Information Theory (ISIT), July 10-15, 2016, Barcelona, Spain. Piscataway: IEEE Press, 2016: 2029-2033.
- [43] ZHANG J, ZHANG X, WANG W. Cache-enabled software defined heterogeneous networks for green and flexible 5G networks[J]. IEEE Access, 2016, 4: 3591-

- 3604.
- [44] SHANMUGAM K, GOLREZAEI N, DIMAKIS G A, et al. FemtoCaching: wireless content delivery through distributed caching helpers[J]. IEEE Transactions on Information Theory, 2013, 59(12): 8402-8413.
- [45] POULARAKIS K, IOSIFIDIS G, TASSIULAS L. Approximation algorithms for mobile data caching in small cell networks[J]. IEEE Transactions on Communications, 2014, 62(10): 3665-3677.
- [46] 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, 2015, 15(8): 1863-1876.
- [47] AVRACHENKOV K, GOSELING J, SERBETCI B. A low-complexity approach to distributed cooperative caching with geographic constraints[C]//The ACM Conference on Measurement and Analysis of Computing Systems (SIGMETRICS), June 5-9, 2017, Urbana-Champaign, USA. New York: ACM Press, 2017.
- [48] LI J, CHEN H, CHEN Y, et al. Pricing and resource allocation via game theory for a small-cell video caching system[J]. IEEE Journal on Selected Areas in Communications, 2016, 34(8): 2115-2129.
- [49] MA G, WANG Z, ZHANG M, et al. Understanding performance of edge content caching for mobile video streaming[J]. IEEE Journal on Selected Areas in Communications, 2017, 35(5): 1076-1089.
- [50] DEHGHAN M, JIANG B, SEETHARAM A, et al. On the complexity of optimal request routing and content caching in heterogeneous cache networks[J]. IEEE/ACM Transactions on Networking (TON), 2017, 25(3): 1635-1648.
- [51] SHNAIWER Y, SOROUR S, SADEGHI P, et al. Network-coded macrocell offloading in femtocaching-assisted cellular networks[J]. IEEE Transactions on Vehicular Technology, 2018, 67(3): 2644-2659.
- [52] HAJRI S E, ASSAAD M. Energy efficiency in cache-enabled small cell networks with adaptive user clustering[J]. IEEE Transactions on Wireless Communications, 2018, 17(2): 955-968.
- [53] GOLREZAEI N, DIMAKIS A G, MOLISCH A F. Scaling behavior for device-to-device communications with distributed caching[J]. IEEE Transactions on Information Theory, 2012, 60(7): 4286-4298.
- [54] JI M, CAIRE G, MOLISCH A F. Wireless device-to-device caching networks: basic principles and system performance[J]. IEEE Journal on Selected Areas in Communications, 2015, 34(1): 176-189.
- [55] YANG C, YAO Y, CHEN Z, et al. Analysis on cache-enabled wireless heterogeneous networks[J]. IEEE Transactions on Wireless Communications, 2016, 15(1): 131-145.
- [56] MADDAH-ALI M A, NIESEN U. Fundamental limits of caching[J]. IEEE Transactions on Information Theory, 2012, 60(5): 2856-2867.
- [57] IOANNIDIS S, MASSOULIE L, CHAINTREAU A. Distributed caching over heterogeneous mobile networks[C]//The ACM Conference on Measurement and Analysis of Computing Systems (SIGMETRICS), June 14-18, 2010, New York, USA. New York: ACM Press, 2010: 311-322.
- [58] QIAO J, HE Y, SHEN X S. Proactive caching for mobile video streaming in millimeter wave 5g networks[J]. IEEE Transactions on Wireless Communications, 2016, 15(10): 7187-7198.
- [59] FILIPPINI I, SCIANCALEPORE V, DEVOTI F, et al. Fast cell discovery in mm-wave 5G networks with context information[J]. IEEE Transactions on Mobile Computing, 2018, 17(7): 1538-1552.
- [60] KELLER L, LE A, CICI B, et al. Microcast: cooperative video streaming on smartphones[C]//The ACM 10th International Conference on Mobile Systems, Applications, and Services (MobiSys), June 25-29, 2012, Ambleside, UK. New York: ACM Press, 2012: 24-25.
- [61] ABDELWAHAB S, HAMDAROU B, GUIZANI M, et al. Replisom: disciplined tiny memory replication for massive IoT devices in LTE edge cloud[J]. IEEE Internet of Things Journal, 2016, 3(3): 327-338.

- [62] NUNNA S, KOUSARIDAS A, IBRAHIM M, et al. Enabling real-time context-aware collaboration through 5G and mobile edge computing[C]//The 12th IEEE International Conference on Information Technology-New Generations (ITNG), April 13-15, 2015, Las Vegas, USA. Piscataway: IEEE Press, 2015.
- [63] CAO W, FENG G, QIN S, et al. Cellular offloading in heterogeneous mobile networks with D2D communication assistance[J]. IEEE Transactions on Vehicular Technology, 2017, 66(5): 4245-4255.
- [64] AO W C, PSOUNIS K. Distributed caching and small cell cooperation for fast content delivery[C]//The 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc), June 22-25, 2015, Hangzhou, China. New York: ACM Press, 2015: 127-136.
- [65] AO W C, PSOUNIS K. Fast content delivery via distributed caching and small cell cooperation[J]. IEEE Transactions on Mobile Computing, 2018, 17(5): 1048-1061.
- [66] KHREISHAH A, CHAKARESKI J, GHARAIBEH A. Joint caching, routing, and channel assignment for collaborative small-cell cellular networks[J]. IEEE Journal on Selected Areas in Communications, 2016, 34(8): 2275-2284.
- [67] YUAN X, WANG X, WANG J, et al. Enabling secure and efficient video delivery through encrypted in-network caching[J]. IEEE Journal on Selected Areas in Communications, 2016, 34(8): 2077-2090.
- [68] KANAI K, MUTO T, KATTO J, et al. Proactive content caching for mobile video utilizing transportation system and evaluation through field experiments[J]. IEEE Journal on Selected Areas in Communications, 2016, 34(8): 2102-2114.
- [69] ZHAO N, LIU X, YU F R, et al. Communications, caching, and computing oriented small cell networks with interference alignment[J]. IEEE Communications Magazine, 2016, 54(9): 29-35.
- [70] ANDREEV S, GALININA O, PYATTAEV A, et al. Exploring synergy between communications, caching, and computing in 5G-grade deployments[J]. IEEE Communications Magazine, 2016, 54(8): 60-69.
- [71] 叶保留, 陈贵海, 郭嵩, 等. 面向云端融合的分布式计算技术研究进展与趋势[R]//CCF 2017-2018中国计算机科学技术发展报告. 2018.
- YE B L, CHEN G H, GUO S, et al. State-of-the-art survey on distributed computing techniques for cloud-edge computing[R]//CCF 2017-2018 annual report on the development of computer science and technology in China. 2018.



## 作者简介



屈志昊 (1989- ), 男, 河海大学计算机与信息学院博士生, 主要研究方向为移动边缘计算、移动流媒体。



叶保留 (1976- ), 男, 博士, 南京大学计算机科学与技术系教授、博士生导师, 中国计算机学会 (CCF) 理事、分布式计算与系统专业委员会秘书长, 主要研究方向为云计算、边缘计算、无线网络。



陈贵海 (1963- ) 男, 博士, 南京大学计算机科学与技术系教授、博士生导师, CCF会士、分布式计算与系统专业委员会主任, 主要研究方向为云计算、边缘计算、计算机网络。



唐斌 (1986- ), 男, 博士, 南京大学计算机科学与技术系助理研究员, CCF会员, 主要研究方向为网络编码、无线网络、边缘计算。



郭成昊 (1981- ), 男, 博士, 就职于中国电子科技集团公司第二十八研究所, 主要研究方向为边缘计算、分布式系统、机器学习。

收稿日期: 2019-01-18

基金项目: 国家重点研发计划基金项目资助 (No.2018YFB1004704); 国家自然科学基金资助项目 (No.61832005); 江苏省重点研发计划基金资助项目 (No. BE2017152)

**Foundation Items:** National Key Research and Development Program of China (No.2018YFB1004704), The National Natural Science Foundation of China(No.61832005), Provincial Key Research and Development Program of Jiangsu (No. BE2017152)