



专题: 5G+6G

面向 B5G 和 6G 的边缘计算与网络切片资源管理

孙茜^{1,2,3}, 田霖^{1,2,3}, 周一青^{1,2,3}, 冯晨^{1,2,3}, 王园园^{1,2,3}, 周继华⁴

(1. 移动计算与新型终端北京市重点实验室, 北京 100190;

2. 中国科学院计算技术研究所无线通信技术研究, 北京 100190;

3. 中国科学院大学, 北京 101408; 4. 重庆金美通信有限责任公司, 重庆 400030)

摘要: 随着时延敏感的计算密集型任务呈指数级增长, 5G 以及未来 6G 的移动蜂窝网络将构建面向边缘计算的切片网络为时延敏感、计算密集型应用提供服务。由于切片网络间共享系统资源, 无线与边缘计算资源的高效管理至关重要。对切片网络生命周期中资源管理的需求进行了研究, 针对资源需求归纳了切片网络接入控制技术、切片网络资源分配技术、切片网络终端协作激励技术的研究进展, 并给出了这 3 方面技术的未来研究方向, 实现了面向边缘计算的切片网络资源的高效管理。

关键词: 网络切片; 边缘计算; 资源管理

中图分类号: TN929.5

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2021128

Resource management in sliced networks with mobile edge computing for B5G and 6G

SUN Qian^{1,2,3}, TIAN Lin^{1,2,3}, ZHOU Yiqing^{1,2,3}, FENG Chen^{1,2,3}, WANG Yuanyuan^{1,2,3}, ZHOU Jihua⁴

1. Beijing Key Laboratory of Mobile Computing and Pervasive Device, Beijing 100190, China

2. Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China

3. University of Chinese Academy of Sciences, Beijing 101408, China

4. Chongqing Jinmei Communication Co., Ltd., Chongqing 400030, China

Abstract: With the time-delay sensitive compute-intensive tasks growing exponentially, 5G and future 6G mobile cellular communication networks will form differentiated network slices with edge computing resource to service the multiple communication services. To achieve this, the wireless frequency and edge computing resources will be shared among the network slices, thus the efficient management of the resources is of great importance. The demands of resources management in the life duration of the network slice was analyzed. Then the research progress of the corresponding technologies for slices, i.e., access control, resources allocation and stimulating cooperation among terminals, were summarized. And the open research issue was given. Thus, the resources are managed efficiently for the slices with edge computing resource.

Key words: network slice, edge computing, resource management

收稿日期: 2021-05-10; 修回日期: 2021-06-10

基金项目: 国家重点研发计划项目 (No.2020YFB1806702); 钱江实验室开放基金资助项目 (No. 2020Y9A003)

Foundation Items: The National Key Research and Development Program of China (No.2020YFB1806702), Open Foundation of Qianjiang Laboratory (No.2020Y9A003)



1 引言

随着以自动驾驶为代表的车联网业务、以虚拟现实/增强现实为代表的视频业务、以远程医疗为代表的智慧生活业务等的涌现,5G 以及未来 6G 时代中时延敏感的计算密集型任务将呈指数级增长。由于计算能力有限的移动设备无法完成上述低时延应用的计算需求^[1],因此将计算任务卸载到边缘进行处理^[2]。移动边缘计算(mobile edge computing, MEC)利用位于移动通信网络边缘的计算资源,对复杂任务进行分布式处理,以提升网络数据处理能力,提供低时延、高容量的计算服务^[3]。未来 5G、B5G(beyond 5G, 后 5G)甚至 6G 网络将利用网络切片为差异化的通信业务提供服务,具体是:基于个性化服务需求为不同垂直行业、不同客户、不同业务提供能力可定制的、相互隔离的网络服务,是一个提供特定网络能力和特性的逻辑网络^[4,5],如图 1 所示。未来 5G/6G 的网络切片将共享无线与边缘计算资源为时延敏感、计算密集型应用提供服务,因此面向网络切片的无线与计算资源管理是关键问题之一。

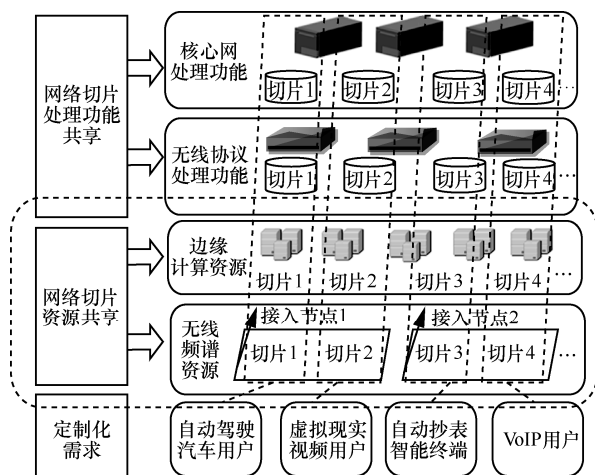


图 1 网络切片资源共享示意图

2 网络切片生命的资源管理需求

网络切片在蜂窝网络软件定义与虚拟化的基

础上实现了整个网络服务定制化。网络切片作为提供服务的方式可以应用于多种垂直行业,根据应用场景、业务类型按需提供网络能力,切片间相互隔离、互不干扰。为此,网络切片运营商与用户之间商定服务等级协议,在服务等级协议(service level agreement, SLA)中包括基本属性(如安全性、可管理性、可用性等)、详细的业务属性(如切片类型、空口参数、差异化网络功能等)以及性能要求(如时延、吞吐率、丢包率等),通过蜂窝网络的虚拟化与物理功能的差异化剪裁与配置有效保证网络切片的 SLA。根据第三方服务提供商/运营商的需求,建立相对应的网络切片,通过网络切片管理器对生成的网络切片进行管理。网络切片管理器从逻辑上为网络切片分配所需的基站功能模块、核心网功能模块、无线与边缘计算资源等。网络切片动态地共享虚拟和/或物理功能模块,根据需求特点对功能模块的能力进行定制化配置,根据网络切片资源分配算法将无线频谱与边缘计算资源合理地分配给相应的切片,网络切片的资源虚拟化架构示意图如图 2 所示。

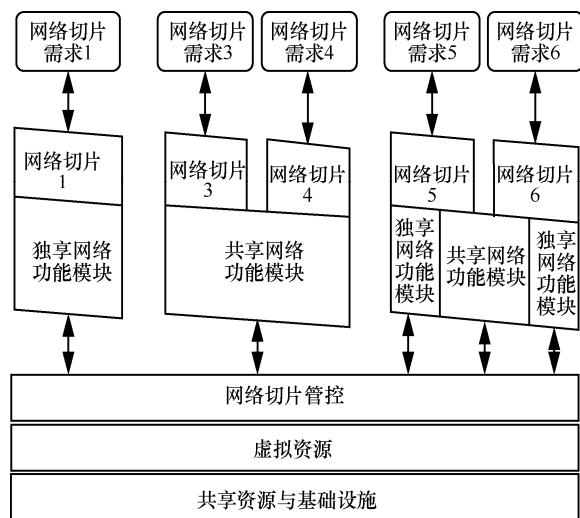


图 2 网络切片的资源虚拟化架构示意图

网络切片的生命周期包含网络切片接入与生成、网络切片运行和网络切片终止 3 个阶段^[6-7],

其中对资源管理的需求具体如下。

(1) 网络切片接入与生成阶段

根据业务需求得到网络切片的资源需求, 运营商判断是否有足够的空余资源构建该网络切片: 如果有, 则接入网络切片; 否则拒绝接入该网络切片。运营商维护和更新一个网络切片模板库, 新业务上线的第一步是匹配相适应的切片模板, 匹配项包括虚拟网络功能组件、组件间标准化交互接口和所需网络资源的描述。当决定接入某网络切片时, 运营商生成该网络切片, 即进行实例化。网络切片实例化时服务引擎导入模板并解析, 通过接口向基础设施提供商租用网络资源, 基于业务需求实例化 VNF 并进行服务功能链的生成与编排, 最后将网络切片迁移到运行态。其中涉及的资源管理为: 根据业务需求得到网络切片的资源需求, 系统为基于模板生成网络切片并基于资源需求预分配资源, 网络切片的资源管理需要能够得到网络切片所需的无线与计算资源需求的平均数值, 并据此完成网络切片资源的预分配。

(2) 网络切片运行阶段

系统对切片的运行状态进行监控、更新、迁移、扩/扩容等操作, 此外系统还支持根据业务负载变化进行快速业务重部署和资源重分配。对于无线与边缘计算资源, 系统需要提供高效的动态资源分配。对于无线协议处理资源, 需要在物理资源池上切分出多个逻辑的资源集合。

(3) 网络切片终止阶段

主要涉及业务下线时功能的去实例化和资源的回收, 以及对资源进行评级、生成历史记录等操作, 一个网络切片的终止不能影响其他切片业务的正常服务。

在网络切片生命周期的资源管理中有 3 个需求, 具体为: (1) 网络切片的接入与生成阶段, 系统通过判断是否有足够的空余资源构建网络切片决策是否接入该切片, 因此需要预估网络切片

的平均无线与计算资源需求; (2) 网络切片运行阶段, 需要系统为网络切片、网络切片为终端动态高效地分配无线与计算资源; (3) 网络切片终止阶段, 由于边缘计算业务需要先通过无线传输给边缘计算服务器后再进行计算处理, 因此无线传输性能将制约边缘计算性能乃至网络切片的整体性能, 面向网络切片需要设计鼓励网络切片与终端参与协作的激励机制。针对网络切片的生命周期的资源管理中的 3 个需求, 后文总结了当前已经开展的网络切片接入控制技术、网络切片资源分配技术、网络切片激励终端协作技术。

3 网络切片接入控制技术

在网络切片接入与生成阶段, 系统需要网络切片的平均无线资源和边缘计算资源需求, 以便于判断是否有足够的空余资源生成该网络切片^[3-4]。若系统有足够的空余资源满足该网络切片的平均资源需求, 则系统接入该网络切片, 并按照需求生成该网络切片, 否则系统拒绝接入该网络切片。针对网络切片仅提供无线传输服务的情况, 现有研究提出了许多方案研究网络切片的无线资源平均需求与无线频谱资源的预设定。

为了支持更丰富的服务质量 (quality of service, QoS), 网络切片可规定多样化的资源需求。对于网络切片的无线资源需求, 具有代表性且研究中广泛应用的是无线传输速率和无线资源量^[4-6]。当采用无线传输速率作为无线资源需求时, 规定网络切片无线传输速率需求的具体数值, 系统根据相应的无线传输速率值为其提供无线资源, 此种方式配置的网络切片适用于有固定无线传输速率要求的业务 (如 IP 语音传输等); 当采用无线资源量作为无线资源需求时, 规定网络切片所需无线资源量的绝对值 (如 5 Mbit/s 带宽) 或者相对值 (如系统带宽的 30%), 系统据此将累计满足需求量的无线资源动态或者固定地分配给网络切片, 此种方式配置的网络切片适用于构建虚拟网



络,即虚拟网络运营商利用分配到的网络资源构建虚拟网络,同时提供语音、数据、视频等多类型的通信业务服务。对于网络切片的计算资源需求,目前还没有相关研究,但在现有面向终端用户的边缘计算资源分配研究中,一般会给定需要处理的数据量/计算量,数据是固定、按照某种规律达到的^[3,8-10]或者是突发的^[11],面向处理数据量/计算量,实时在线分配无线与计算资源。

上述网络切片的无线资源需求未考虑与计算资源的耦合关系,边缘计算服务需求是从业务的角度在线实时调整,无法提供资源需求的具体数值。为了解决这一问题,提出了基于多队列模型研究无线传输速率波动情况下网络切片的平均无线与计算资源需求的耦合与置换关系。如图3所示,假设每个网络切片有两个先入先出(first input first output, FIFO)串联队列,分别为无线传输队列和MEC计算队列,网络切片的待卸载边缘计算任务在无线传输队列中排队传输给AP,AP直接通过有线方式传输给MEC服务器,计算任务在MEC计算队列中排队等待计算处理。

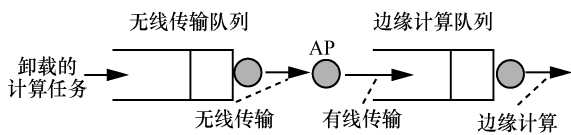


图3 网络切片边缘计算卸载任务的队列模型

所以在给定边缘计算任务,并给出了网络切片所需的平均无线与计算资源时,以最大化系统接入网络切片数为目标,可以建模出网络切片的接入控制问题,以实现系统资源的高效利用,提升系统对网络切片的接入能力^[12]。无业务突发场景JMCR算法与对比算法的切片总数与系统接入网络切片数的关系如图4所示,将该文提出的JMCR算法和Com1、Com2算法进行了对比,显然引入了无线与计算资源的置换关系的JMCR算法可以服务更多的网络切片,使系统获得了更大的接入网络切片数目。

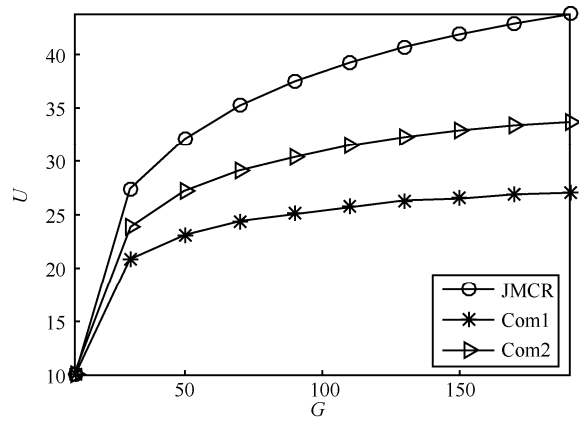


图4 无业务突发场景JMCR算法与对比算法的切片总数与系统接入网络切片数的关系

4 网络切片资源分配技术

未来移动通信网络将构建网络切片提供数据无线传输与边缘计算服务,在网络切片运行的过程中,无线与计算资源由系统先分配给网络切片再被网络切片中的终端共享,因此有效分配无线与计算资源很重要。

4.1 网络切片无线资源分配

现有网络切片资源管理架构范式为网络虚拟基片(network virtualization substrate, NVS),由切片层与终端层构成,其中切片层完成面向网络切片的资源分配,终端层完成各网络切片内部面向终端用户的资源分配^[13]。因为面向终端层的资源分配可采取传统非切片网络中的方法(如比例公平算法、轮询算法等),大多数研究关注切片层的资源分配,设计满足网络切片资源需求的资源分配方法。

为此,当前的网络切片无线资源分配技术从参数设定和资源分配算法设计两个方面展开。针对网络切片无线资源分配参数设定,首先在LTE网络中研究了基于预留的切片层无线资源分配机制,当新切片接入时若空闲预留资源或剩余资源能够满足其需求则接入并其按需求分配资源,否则拒绝^[14]。进而,研究了面向多网络切片的接入网协议层二和层三的差异化参数配置,通过个性

化配置网络切片资源使用参数配置和可用资源参数配置实现差异化网络切片服务质量保障, 阐明网络切片无线资源分配参数配置对网络切片的服务能力定制化与服务质量保障的影响^[15]。针对资源分配算法设计, 首先基于对网络切片的资源利用率, 利用线性规划方法对网络切片进行动态资源调整, 低资源利用率的网络切片向高资源利用率的切片提供资源^[16]。然后, 面向超可靠低时延通信 (ultra-reliable and low-latency communication, uRLLC) 切片、IoT (internet of thing, 物联网) 切片和增强型移动宽带 (enhanced mobile broadband, eMBB) 切片的差异化 QoS 要求, 研究了网络切片间和切片内的无线资源分配算法, 实现了网络整体利用率的提高^[17]。同时, 针对云接入网研究了基于终端用户和切片的双优先级的无线资源分配机制, 以实现更高的网络吞吐量和 QoS 性能^[18]。此外, 针对雾接入网, 基于匈牙利算法、线性整数规划以及几何规划提出了网络资源缓存策略以减小网络切片时延以及优化无线资源动态分配^[19]。

上述研究大多基于优化理论, 此外面向网络切片的无线资源管理可基于机器学习^[20]、博弈论^[21]、拍卖模型^[22]、广义速率调度^[23]、合同理论^[24]和复杂网络理论^[25]等开展。

当前的网络切片无线资源分配独立进行而未考虑与计算资源的耦合关系, 而面向边缘计算的切片需要利用无线资源将计算业务传输到边缘服务器后再进行计算处理, 因此, 无线资源与计算资源的分配需要联合考虑。

4.2 网络切片计算资源分配

虽然目前没有针对网络切片的边缘计算资源分配研究, 但针对无切片网络的边缘计算分配已经开展了大量的研究。面向单小区多终端用户场景^[8]、多小区单终端用户场景^[9]、多小区多终端用户场景^[10], 通过卸载决策和计算资源分配最小化时延和/或能源消耗^[26]。为获得更短的平均执行时

延, 允许在移动设备和边缘计算服务器上并行执行计算任务, 移动设备的计算任务可跨越多无线信道传输、跨多个边缘计算服务器处理, 并提出了一种基于马尔可夫决策方法的计算任务调度和边缘计算资源分配策略^[27]。考虑系统的能源消耗的约束, 结合负载均衡思想, 研究了边缘节点协同处理计算任务, 同时有效降低计算任务在边缘服务器中的排队时延, 从而提升边缘计算性能^[28]。上述研究中仅涉及计算资源分配, 没有考虑无线资源分配。

为此, 针对边缘计算与无线资源联合分配开展了进一步研究^[8,11,29-31]。考虑无线和计算资源, 基于移动网络的流量分布不均匀的特征, 提出了一种基于睡眠控制的在线优化策略, 以解决 MEC 服务器计算任务卸载的问题。并采用 Lyapunov 技术进一步优化了该问题的求解过程, 从而在降低 MEC 服务器网络的长期能耗的同时也可以保持较低的系统时延^[11]。在多小区场景中, 根据终端的支付能力或工作类型为其设置卸载优先级, 建模了具有固定任务卸载决策的资源分配问题, 利用凸优化技术对此类混合整数非线性规划问题进行求解, 实现了系统效益最大化^[29]。在 MIMO (multi-input multi-output, 多输入多输出) 多小区场景中, 系统将无线和计算资源的联合分配建模成优化问题, 提出基于云管理器与基站交换信息的最优无线和计算资源分配方案, 在单终端用户场景下计算出封闭形式的全局最优解, 在多终端用户场景下采用基于服务组件架构的集中式和分布式算法得到局部最优解^[30]。为防止传输干扰和 MEC 服务器过载, 选择部分移动终端卸载其计算, 其他终端则在本地执行计算, 根据终端不同的需求分配相应的计算和无线资源, 通过采用基于交替方向乘子法的优化算法解决计算卸载、计算和无线资源分配及内容缓存问题^[31]。上述研究中的计算与无线资源分配为静态的, 根据初始的资源请求为终端分配固定量的计算与无线资源。



静态机制的无线与计算资源分配方法虽然复杂度低,但灵活性差、资源效率与用户体验差。

针对静态机制的计算与无线联合资源分配存在的问题,对动态机制的计算与无线资源联合分配展开了研究,根据终端/业务的动态资源需求实时按需分配资源^[32-35]。在要求密集计算、实时响应但又资源受限的场景中,提出了联合无线与计算资源分配的框架,基于此框架研究了在 QoS 约束下的卸载决策和在最小资源占用约束条件下的无线与计算资源联合分配,可以满足不同移动终端的丰富需求^[32]。针对小蜂窝网络场景,由于多终端同时卸载计算任务将带来无线干扰,降低无线传输速率,为此研究计算卸载和干扰管理问题,根据终端计算开销和卸载开销的估计值作出卸载决策,通过图着色进行物理资源块 (physical resource block, PRB) 分配,依据决策结果和 PRB 分配结果分配计算资源^[33]。为平衡无线与计算资源的使用,提出一种基于深度强化学习的智能无线与计算资源分配方案,可在不同 MEC 条件下改进资源分配策略,实现资源的自适应分配^[34]。在移动设备能力有限但其计算需求不断增长的场景中,提出由传输队列和计算处理队列组成的两级串联队列建模无线传送与计算处理,推导 MEC 服务器的有效容量,分析在端到端时延方面无线和计算资源之间的内在关系,并制定了无线和计算资源的联合分配方案以使网络总收益最大化,进而采用交替方向乘数法 (alternating direction method of multipliers, ADMM) 对无线与计算资源联合分配方案进行求解^[35]。动态机制的计算与无线资源联合分配方法虽然灵活且资源效率高、用户体验好,但是算法的复杂度较高。

为了解决动态机制的资源分配方法复杂度高的问题,参考文献[36-37]对低复杂度的资源分配方法开展研究。在无弹性计算任务的场景中,提出了一种多终端用户的无线资源和计算资源分配方案,设计不可忽略任务执行时间的计算模型,

建模系统消耗能源最小的多终端计算与无线资源联合分配问题,并利用 Johnson 算法求解,算法复杂度为计算任务队列的平方到三次方之间^[36]。针对集中化无线电接入网 (centralized RAN, C-RAN) 场景,设计可以动态联合调度无线资源和 MEC 计算资源的框架,基于扩展的 Lyapunov 技术设计出作业可变长度算法,在保持系统稳定的同时达到接近最优的平均时间收益,计算复杂度接近终端数目的平方^[37]。针对冗余性和动态性的计算业务,研究联合有线和无线链路上 MEC 选择和路由优化,提出一种基于博弈论的低复杂度的资源分配算法,最小化 MEC 网络的能量消耗和平均响应时间,鉴于实际情况下重路由带来的成本,又提出一种基于 β 改进偏差的快速收敛算法,通过建立屏障减小重路由收敛速度提高 30% 以上^[38]。上述研究中虽然降低了动态机制的计算与无线资源联合分配算法的复杂度,但无法直接应用在网络切片的场景中。现有面向终端展开的研究是单层的无线与计算的资源分配,网络切片场景中系统需要将资源先分配给切片层再分配给终端层,由于切片层的资源需求受到终端层的影响,两层的资源分配需要联合进行。

现有的基于静态机制的无线与计算资源联合分配方法虽然复杂度较低但不灵活,这将导致资源浪费或用户体验差;现有的基于动态机制的无线与计算资源联合分配方法高效且灵活,但具有很高的复杂性。现有低复杂度的单层资源分配动态机制,无法直接应用到网络切片场景中。因此面向网络切片研究低复杂度的无线与计算资源分配,针对网络切片资源需求为其服务的多终端资源需求的叠加这一特征,设计了半动态的无线与计算资源分配机制,即 SD-CRA。即在每个分配周期期间,分配给网络切片的无线与边缘计算资源动态调整。网络切片共享 MEC 资源示意图如图 5 所示,其中虚拟 MEC (virtual MEC, VMEC),表示用于一个网络切片的 MEC 计算资

源,可能由一个或多个 MEC 服务器的部分或全部计算资源组成。

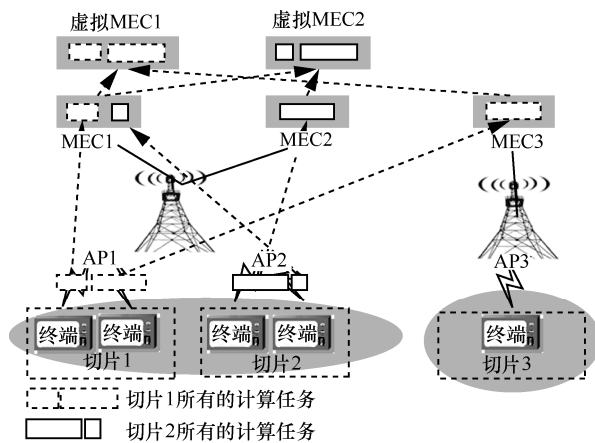


图5 网络切片共享 MEC 资源示意图

在 SD-CRA 机制中,仅在更新周期内系统会更新分配给网络切片的无线与计算资源,在 SD-CRA 机制周期内无线与计算资源静态地分配给网络切片,同时在 SD-CRA 机制更新周期内网络切片服务的终端用户动态共享网络切片分配到的无线与计算资源,实现了面向网络切片的低复杂度与高效率的无线与计算资源分配^[39]。网络切片 VMEC 负载率与数据量之间的关系如图 6 所示,在复杂性和效率方面,将提出的 SD-CRA 机制算法与一种静态算法 STA 和一种动态算法 DYA 进行了比较。其中, WCE 表示信道容量非受限下的无线与计算资源联合分配算法, MCA 表示信道容量受限下以最小化 VMEC 的计算资源为目标的资源分配算法, MTA 表示信道容量受限下以最大化 VMEC 的总无线吞吐量为目标的资源分配算法。可以看出 SD-CRA 机制算法比 STA 的效率(负载率)要高,比 DYA 的复杂性低。

5 网络切片激励终端协作技术

5G 移动通信网络通过构建网络切片提供差异化的服务,由于视频回传、远程控制等多种场景需要边缘计算,5G 网络切片将包含无线与计算资源。因为边缘计算业务需要先通过无线传输到

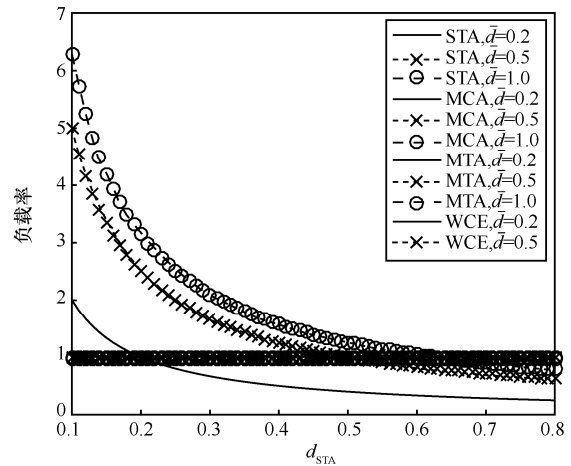


图6 网络切片 VMEC 负载率与数据量之间的关系

边缘计算服务器再进行计算处理,因此无线传输能力将通过制约计算能力影响网络切片的服务能力。为此网络切片无线传输能力提升技术的研究十分重要。通过网络切片间的终端用户协作,即通过某一网络切片的终端用户为另一网络切片的终端用户做 D2D 中继,提升部分终端用户的无线传输能力,进而提升网络切片的无线传输能力。在网络切片间终端用户 D2D 协作中,需要在网络切片间、网络切片与其服务的终端用户间、协作与被协作的终端用户间交互控制和/或数据信息,而考虑网络切片与终端用户的自私性,需要设计激励机制,鼓励网络切片与终端用户提供协作。

虽然没有针对基于网络切片的终端协作激励的研究,但是针对无切片网络的终端协作激励已经展开了大量的研究,主要分为从用户终端角度出发进行研究和从整体网络角度出发进行研究。从用户终端角度出发的包括:以提高协作与被协作终端用户的吞吐量为目标,针对无线信道质量较差的终端用户,提出的一种激励性资源分配方案,通过引入能量效率作为激励参数来激励终端用户之间的协作,设计了一种二维搜索方法解决双终端用户案例问题^[40];以终端用户体验质量最大化为目标,针对长期演进(long term evolution, LTE)的演进(LTE-advanced, LTE-A)网络中的



D2D 通信, 提出的一种基于资源交换的协作激励内容分发策略, 根据网络拓扑结构和终端用户的意见评分构造加权有向图, 由于每个节点都要向其他节点发送内容以获得所需的内容, 从而激励终端用户相互协作的共享内容, 增强系统的鲁棒性^[41]。从整体网络角度出发的包括: 以最大限度地提高移动运营商的效用为目标, 采用 D2D 协作数据卸载技术, 提出的一种集中式缓存的激励机制, 激励终端参与协作数据的流量加载, 在蜂窝网络中基于反向拍卖理论建模缓存终端选择和内容分配问题, 根据终端将租用的存储空间大小和承诺的协作行为选择合适的终端^[42]; 以实现网络能量效率最大化为目标, 为降低在 D2D 多播内容共享场景中的传输能耗, 提出的一种基于社会关系信息的激励机制, 即通过社会关系网对终端进行聚类, 作为簇头的终端用户将承担计算任务, 而其他终端用户则根据相应的社会关系信息义务性或有回报地帮助簇头将内容传送到目标终端用户^[43]。另外为防止资源被分配给没有为协作做出贡献的终端用户, 参考文献^[44]提出一种基于虚拟货币系统的激励方案, 为每一个终端用户设置虚拟货币账户, 货币可以用于购买其他终端用户的协作服务, 通过采用两级 Stackelberg 博弈模型, 考虑内容请求者、协作中继终端用户和内容源的收益。

以上这些激励机制通常是通过为终端用户设置虚拟标签 (如虚拟现金和声誉) 来反映中继终端用户的协作利益, 系统将根据终端用户的虚拟标签来分配资源, 以激励终端用户主动为其他终端用户提供协作^[40-41, 44-45], 部分激励机制利用社会关系信息鼓励终端用户参与合作, 利用社会学中的同质现象, 终端用户之间根据亲密性互相帮助传送内容^[41-46]。因此, 理性的终端用户会考虑他们未来被协作的好处, 同时会得到与其他终端用户协作的奖励^[40, 47-48]。然而现有的 D2D 协作通信激励方案不能直接推广到切片网络中的协作, 因

为在双层体系结构中, 首先将资源分配给切片, 然后在其所服务的终端用户之间进行分配, 因此作为奖赏而分配给中继移动终端 (mobile terminal, MT) 用户的资源应该提前分配给切片。因此面向网络切片需要设计双层的激励 D2D 协作的资源分配方法, 将提出其建模为两个无线资源分配的最大化效用函数问题, 在优化问题的效用函数中分别引入了网络切片和终端用户的协作行为。其以下行数据传输场景为主要研究目标, 一个 AP 服务 G 个切片示意图如图 7 所示, 数据由共享 AP 通过下行无线链路发送给移动终端用户, 并假设无线频谱资源以 TDMA 模式共享。

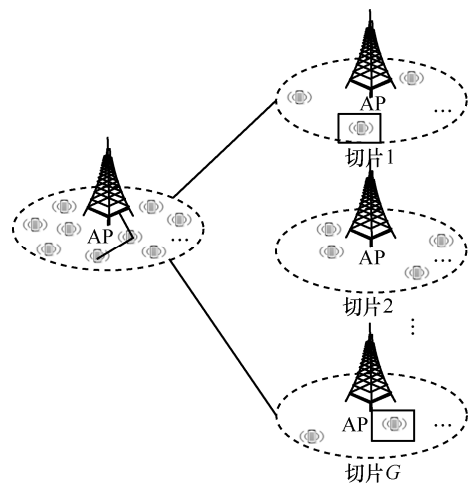


图 7 一个 AP 服务 G 个切片示意图

当为提供协作的切片与终端分配更多的无线资源时, 效用函数会倍增, 从而将更多的资源作为奖励分配给协作的切片和终端用户, 由于网络切片和终端用户的瞬时资源分配不能通过两个优化问题得到, 进而设计了网络切片和终端用户的激励调度算法, 并定义隔离度指标以评估执行协作时的网络切片的隔离性能^[49-50]。如图 8 所示, 可以看到隔离度在系统以非协作模式 (non-cooperative mode, NCM) 运行时中较差, 而在系统以切片协作模式 (slice cooperation mode, SCM) 运行中得到了改进。

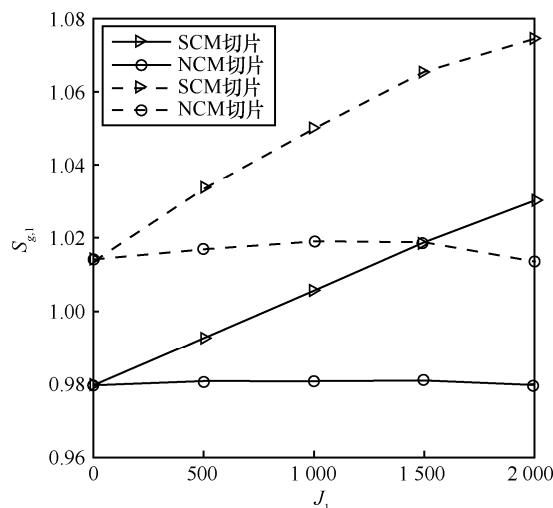


图8 网络切片隔离指数

6 开放研究方向

(1) 网络切片的资源隔离性能与资源复用之间的折中关系研究

5G 以及 B5G 通过构建网络切片提供差异化的服务, 由于网络切片间共享的资源与基础设施, 通过资源复用, 可有效降低建设与运维成本, 提供网络的灵活性。与此同时, 为保障网络切片的服务质量, 网络切片间资源的隔离是一个重要的问题。然而网络切片的资源复用与隔离是一对相互矛盾的性能指标, 如何建模多切片系统的复用性能和隔离性能、推导两者之间的制约机理是一个复杂的问题。可对多切片系统的复用性能和隔离性能进一步建模, 分析推导多切片系统的复用性能和隔离性能折中机理, 这对多切片系统的资源高效管理、服务质量保障具有重要意义。

(2) 基于隔离性能的网络切片资源管理研究

由于隔离性能是网络切片的关键性能, 且隔离性能中包含了网络切片中切片层与终端层的各类性能参数的满足情况, 如何保障网络的隔离性能是网络切片的重要问题。而当前网络切片资源管理方法研究中几乎没有考虑网络切片隔离性能的量化度量以及网络切片隔离性能保障的资源管理, 而网络切片的隔离性能由于涉及切片层与终

端层双层以及多类别的参数, 如何建模隔离性能参数与网络切片双层资源管理的动态调节关系是一个复杂的问题, 可继续深入开展研究。

7 结束语

为服务时延敏感的计算密集型业务, 移动蜂窝网络将构建面向边缘计算的网路切片。由于网络切片间共享系统资源, 无线与边缘计算资源的高效管理至关重要。本文分析了网络切片生命周期中资源管理的需求, 针对网络切片资源管理的需求进一步归纳了相应的研究进展, 并进一步给出了未来研究方向, 这对多切片系统的资源高效管理、服务质量保障具有重要意义。

参考文献:

- [1] PEREZ M, XU S, CHAUHAN S, et al. Impact of delay on telesurgical performance: study on the robotic simulator dV-Trainer[J]. International Journal of Computer Assisted Radiology & Surgery, 2016, 11(4): 581-587.
- [2] PAWANI P, JUDE O, MADHUSANKA L, et al. Survey on multi-access edge computing for Internet of Things realization[J]. IEEE Communications Surveys & Tutorials, 2018, 20(4): 2961-2991.
- [3] MAO Y, YOU C, ZHANG J, et al. A survey on mobile edge computing: the communication perspective[J]. IEEE Communications Surveys & Tutorials, 2017, 19(4): 2322-2358.
- [4] 3GPP. Study on new radio access technology: radio access architecture and interfaces V14.0.0: TR 38.801[S]. 2017.
- [5] GILBERT T. \$1.4tn of benefits in 2030: 5G's impact on industry verticals [EB]. 2019.
- [6] KALOXYLOS. A survey and an analysis of network slicing in 5G networks[J]. IEEE Communications Standards Magazine, 2018, 2(1): 60-65.
- [7] ZHANG S. An overview of network slicing for 5G[J]. IEEE Wireless Communications, 2019, 26(3): 111-117.
- [8] 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, 2017, 17(3): 1784-1797.
- [9] DINH T Q, TANG J, LA Q D, et al. Offloading in mobile edge computing: task allocation and computational frequency scaling[J]. IEEE Transactions on Communications, 2017, 65(8): 3571-3584.
- [10] CHEN M, HAO Y. Task offloading for mobile edge computing



- in software defined ultra-dense network[J]. *IEEE Journal on Selected Areas in Communications*, 2018, 36(3): 587-597.
- [11] WANG S, ZHANG X, YAN Z, et al. Cooperative edge computing with sleep control under non-uniform traffic in mobile edge networks[J]. *IEEE Internet of Things Journal*, 2018, 6(3): 4295-4306.
- [12] SUN Q, TIAN L, SHI J, et al. Joint management of communicating and computing resources in sliced 5G networks[C]//*Proceedings of GLOBECOM 2020 - 2020 IEEE Global Communications Conference*. Piscataway: IEEE Press, 2020: 1-6.
- [13] KOKKU R, MAHINDRA R, ZHANG H, et al. NVS: a substrate for virtualizing wireless resources in cellular networks[J]. *IEEE/ACM Transactions on Networking*, 2012, 20(5): 1333-1346.
- [14] GUO T, ARNOTT R. Active LTE RAN sharing with partial resource reservation[C]//*Proceedings of 2013 IEEE 78th Vehicular Technology Conference (VTC Fall)*. Piscataway: IEEE Press, 2013: 1-5.
- [15] FERRUS R, SALLENT O, PEREZ-ROMERO J, et al. On 5G radio access network slicing: radio interface protocol features and configuration[J]. *IEEE Communications Magazine*, 2018, 56(5): 184-192.
- [16] CHEN J L, MA Y W, KUO H Y, et al. Enterprise visor: a Software-Defined enterprise network resource management engine[C]//*Proceedings of 2014 IEEE/SICE International Symposium on System Integration*. Piscataway: IEEE Press, 2014: 381-384.
- [17] ZHANG H, LIU N, CHU X, et al. Network slicing based 5G and future mobile networks: mobility, resource management, and challenges[J]. *IEEE Communications Magazine*, 2017, 55(8): 138-145.
- [18] LEE Y L, LOO J, CHUAH T C, et al. Dynamic network slicing for multitenant heterogeneous cloud radio access networks[J]. *IEEE Transactions on Wireless Communications*, 2018, 17(4): 2146-2161.
- [19] TANG L, ZHANG X, XIANG H, et al. Joint resource allocation and caching placement for network slicing in fog radio access networks[C]//*Proceedings of 2017 IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. Piscataway: IEEE Press, 2017: 1-6.
- [20] XU Q, WANG J, WU K. Learning-based dynamic resource provisioning for network slicing with ensured end-to-end performance bound[J]. *IEEE Transactions on Network Science & Engineering*, 2018, 7(1): 28-41.
- [21] KAZMI S M A, TRAN N H, HO T M, et al. Hierarchical matching game for service selection and resource purchasing in wireless network virtualization[J]. *IEEE Communications Letters*, 2018, 22(1): 121-124.
- [22] JIANG M, CONDOLUCI M, MAHMOODI T. Network slicing in 5G: an auction-based model[C]//*Proceedings of 2017 IEEE International Conference on Communications (ICC)*. Piscataway: IEEE Press, 2017: 1-6.
- [23] MALANCHINI I, VALENTIN S, AYDIN O. Generalized resource sharing for multiple operators in cellular wireless networks[C]//*Proceedings of 2014 International Wireless Communications and Mobile Computing Conference (IWCMC)*. Piscataway: IEEE Press, 2014: 803-808.
- [24] NGUYEN D H N, YANRU Z, ZHU H. Contract-based spectrum allocation for wireless virtualized networks[J]. *IEEE Transactions on Wireless Communications*, 2018, 17(11): 7222-7235.
- [25] SHEN Y, WEN X, WANG L, et al. Physical resource management based on complex network theory in 5G network slice trading[C]//*Proceedings of 2019 IEEE Wireless Communications and Networking Conference (WCNC)*. Piscataway: IEEE Press, 2019: 1-6.
- [26] MACH P, BECVAR Z. Mobile edge computing: a survey on architecture and computation offloading[J]. *IEEE Communications Surveys & Tutorials*, 2017, 19(3): 1628-1656.
- [27] LIU J, MAO Y, ZHANG J, et al. Delay-optimal computation task scheduling for mobile-edge computing systems[C]//*Proceedings of 2016 IEEE International Symposium on Information Theory (ISIT)*. Piscataway: IEEE Press, 2016: 1451-1455.
- [28] JIA M, CAO J, LIANG W. Optimal cloudlet placement and user to cloudlet allocation in wireless metropolitan area networks[J]. *IEEE Transactions on Cloud Computing*, 2017, 5(4): 725-737.
- [29] TRAN T X, POMPILI D. Joint task offloading and resource allocation for multi-server mobile-edge computing networks[J]. *IEEE Transactions on Vehicular Technology*, 2019, 68(1): 856-868.
- [30] SARDELLIOTTI S, SCUTARI G, BARBAROSSA S. Joint optimization of radio and computational resources for multicell mobile-edge computing[J]. *IEEE Transactions on Signal & Information Processing over Networks*, 2015, 1(2): 89-103.
- [31] WANG C, LIANG C, YUV F R, et al. Joint computation offloading, resource allocation and content caching in cellular networks with mobile edge computing[C]//*Proceedings of 2017 IEEE International Conference on Communications (ICC)*. Piscataway: IEEE Press, 2017: 1-6.
- [32] XU C, WENZHONG L, SANGU L, et al. Efficient resource allocation for on-demand mobile-edge cloud computing[J]. *IEEE Transactions on Vehicular Technology*, 2018, 67(9): 8769-8780.
- [33] WANG C, YU F, LIANG C, et al. Joint computation offloading and interference management in wireless cellular networks with mobile edge computing[J]. *IEEE Transactions on Vehicular Technology*, 2017, 66(8): 7432-7445.
- [34] WANG J, ZHAO L, LIU J, et al. Smart resource allocation for mobile edge computing: a deep reinforcement learning approach[J]. *IEEE Transactions on Emerging Topics in Compu-*

- ting, 2019: 1-17.
- [35] WANG Y, TAO X, HOU Y T, et al. Effective capacity-based resource allocation in mobile edge computing with two-stage tandem queues[J]. IEEE Transactions on Communications, 2019, 67(9): 6221-6233.
- [36] GUO J, SONG Z, CUI Y, et al. Energy-efficient resource allocation for multi-user mobile edge computing[C]// Proceedings of GLOBECOM 2017 - 2017 IEEE Global Communications Conference. Piscataway: IEEE Press, 2017: 1-7.
- [37] WANG X, WANG K, WU S, et al. Dynamic resource scheduling in mobile edge cloud with cloud radio access network[J]. IEEE Transactions on Parallel and Distributed Systems, 2018, 29(11): 2429-2445.
- [38] ZHOU Z, MA G, XU C, et al. Energy-efficient resource allocation in cognitive D2D communications: a game-theoretical and matching approach[C]// Proceedings of 2016 IEEE International Conference on Communications (ICC). Piscataway: IEEE Press, 2016: 1-6.
- [39] SUN Q, TIAN L, ZHOU Y, et al. Semi-dynamic computing resource allocation in MEC-enabled radio access networks[C]// Proceedings of 2019 IEEE Global Communications Conference (GLOBECOM). Piscataway: IEEE Press, 2019: 1-6.
- [40] SUN Q, TIAN L, ZHOU Y, et al. Energy efficient incentive resource allocation in D2D cooperative communications[C]// Proceedings of 2015 IEEE International Conference on Communications (ICC). Piscataway: IEEE Press, 2015: 2632-2637.
- [41] ZHANG Q, GUI L, TIAN F, et al. A caching-based incentive mechanism for cooperative data offloading[C]// Proceedings of 2017 IEEE International Conference on Communications Workshops (ICC Workshops). Piscataway: IEEE Press, 2017: 1376-1381.
- [42] FANG Y, ZHOU Y, JIANG X, et al. Game theoretic D2D content sharing: joint participants selection, routing and pricing[C]// Proceedings of 2017 26th International Conference on Computer Communication and Networks (ICCCN). Piscataway: IEEE Press, 2017: 1-9.
- [43] ZHANG A, ZHOU L, WANG L. QoE-driven scheme for multimedia content dissemination in Device-to-Device communications and Mobile Computing Conference (IWCMC). Piscataway: IEEE Press, 2014: 785-790.
- [44] HE Z, ZEHUA W, QINGHE D. Social-aware D2D relay networks for stability enhancement: an optimal stopping approach[J]. IEEE Transactions on Vehicular Technology, 2018, 67(9): 8860-8874.
- [45] DAN W, LIANG Z, YUEMING C, et al. Optimal content sharing mode selection for social-aware D2D communications[J]. IEEE Wireless Communications Letters, 2018, 7(6): 910-913.
- [46] WU D, ZHOU L, CAI Y. Social-aware rate based content sharing mode selection for D2D content sharing scenarios[J]. IEEE Transactions on Multimedia, 2017, 19(11): 2571-2582.
- [47] XU H, XU W, YANG Z, et al. Pilot reuse among D2D users in D2D underlaid massive MIMO systems[J]. IEEE Transactions

on Vehicular Technology, 2018, 67(1): 467-482.

- [48] LIU G, YU F R, JI H, et al. In-band full-duplex relaying: a survey, research issues and challenges[J]. IEEE Communications Surveys & Tutorials, 2017, 17(2): 500-524.
- [49] SUN Q, TIAN L, ZHOU Y, et al. A two-layered incentive scheme for cooperation in sliced 5G D2D networks[J]. IEEE Transactions on Vehicular Technology, 2020, 69(11): 13289-13304.
- [50] SUN Q, TIAN L, ZHOU Y, et al. Incentive scheme for slice cooperation based on D2D communication in 5G networks[J]. China Communications, 2020, 17(1): 28-41.

[作者简介]



孙茜 (1990-), 女, 博士, 中国科学院计算技术研究所助理研究员, 主要研究方向为移动通信和无线计算资源分配。



田霖 (1980-), 女, 博士, 中国科学院计算技术研究所研究员, 主要研究方向为无线资源管理、无线多媒体多播、新型无线网络架构。



周一青 (1975-), 女, 博士, 中国科学院计算技术研究所研究员, 主要研究方向为无线移动通信技术, 包括多天线技术、多载波 OFDM、协同传输、干扰抵消、信道编/解码、网络编码、信息论、绿色无线电等。

冯晨 (1997-), 女, 中国科学院计算技术研究所硕士生, 主要研究方向为移动通信和 5G 专网。

王园园 (1986-), 女, 博士, 中国科学院计算技术研究所高级工程师, 主要研究方向为计算与通信融合无线移动通信系统架构、网络资源管理、运行控制关键技术。

周继华 (1979-), 男, 博士, 重庆金美通信有限责任公司研究员、总经理助理, 主要研究方向为 5G 和 6G 通信、智能集群通信。