
目录

第四部分 MEC 关键技术.....	3
1 技术分类	3
2 服务部署	3
2.1 简介	3
2.2 基于 LTE 网络的部署	4
2.3 基于 5G 网络的部署	5
2.4 基于融合网络的部署	6
2.5 参考文献	8
3 本地分流	8
3.1 简介	8
3.2 关键技术	8
3.3 研究进展	11
3.4 参考文献	12
4 缓存加速	13
4.1 CDN 简介	13
4.2 研究进展	14
4.3 参考文献	16
5 移动性管理	17
5.1 简介	17
5.2 MEC 移动性管理	18
5.3 应用迁移分类	19
5.4 研究进展	20
5.5 参考文献	22
6 网络能力开放	22
6.1 概述	22
6.2 3GPP SCEF	23
6.3 API 标准综述	25
6.4 参考文献	28

7	计算卸载	29
7.1	概述	29
7.2	主要流程	30
7.3	关键技术	31
7.4	研究进展	35
7.5	参考文献	37
8	多接入管理	38
8.1	概述	38
8.2	MAMS	38
8.3	MAMS 与 MEC	40
8.4	参考文献	41
9	资源管理	41
9.1	概述	41
9.2	关键技术	41
9.3	研究进展	43
9.4	参考文献	45
10	安全管理	46
10.1	概述	46
10.2	安全需求	46
10.3	安全框图	48
10.4	关键技术	49
10.5	研究进展	51
10.6	参考文献	52

第四部分 MEC 关键技术

前面章节中，我们分别介绍了 MEC 概述、MEC 标准概要、MEC 系统架构等内容，这是一个从理论到实践的过程。本章将从 MEC 学术和基础技术角度对 MEC 技术进行讲解，为解决各类 MEC 技术问题提供理论和实践基础。MEC 技术研究的领域主要包括：本地分流、缓存加速、移动性管理（包括服务和应用迁移）、网络能力开放、计算卸载、服务部署、协同计算、NFV 融合、固网融合、安全管理、计费策略、资源管理等。最后，从计算机和通信基础技术角度对 MEC 基础技术进行介绍。

1 技术分类

MEC 为应用程序开发者和内容提供者，提供云计算能力和移动边缘网络的 IT 服务环境，从而实现超低时延、高带宽、实时性的网络信息访问。MEC 所有的技术需求都是从业务实际需求出发，其核心是对外提供符合各种业务部署场景下应用，并创造价值，主要包括如下关键技术：本地分流、缓存加速、移动性管理、计算卸载、服务部署、协同计算等，都是基于 MEC 业务服务角度产生的技术需求。参考如下表，对 MEC 技术分为：接口相关、架构、安全、业务服务、容器化/虚拟化、应用和部署、开发工具。

类名	说明
接口	包括云计算中心与 MEC 服务间、MEC 服务与 MEC 服务间、设备（终端、传感器、移动设备等）与 MEC 服务，MEC 服务与其他服务间的交互等；交互涉及到的通信协议、通信规则、策略、标准业务数据格式化等。
架构	包括 MEC 架构、MEC 与 4G、5G、NFV 融合架构、以及 MEC 网络功能服务架构；以及 MEC 相关的与数据中心融合硬件架构、通用 MEC 硬件架构等。
安全	ME 防篡改、安全架构，业务隐私、加密方法、防火墙等。
业务服务	包括 MEC 服务的可靠性，可达性，服务质量，服务迁移，认证，流量计算，收费，负载均衡，关键/实时计算服务、本地分流、CDN、TCP 优化、AI 计算、实时数据分析、移动性支持、服务控制、业务链、接入技术、流量优化、智能采集等。
容器化/虚拟化	包括 SDN、NFV、存储、编程模型、轻量级系统和函数库、数据旁路、微服务、数据库、MANO 等。
应用和部署	包括各类 MEC 应用、业务部署策略、MEC 应用软件开发等，如 IoT、车联网、企业园区、工业自动化、体育馆、VR/AR、无人机、商场、医院、学校等各种场景下的业务应用。
开发工具	标准化应用程序接口的软件开发，包括支持弹性边缘应用的软件 SDK、测试维护、指标量化工具等。

2 服务部署

2.1 简介

MEC 服务部署（或称部署策略：Deployment Strategy）的核心目的是在用户资源需求和约束限制的前提下，达到提高资源利用率，减少网络时延等。目前，MEC 服务部署应该如何放置还没有明确的定义，一般是根据具体的网络环境及业务需求，包括 MEC 主机和边缘云在年内的 MEC 基础设施可部署在移动网络的不同位置。不同的 MEC 应用场景、网络运营商、内容提供商对 MEC 部署等存在不同诉求。特别是 5G 三大场景 eMBB（增强移动宽带）、URLLC（极可靠低延迟通信）和 mMTC（海量机器通讯），合理有效的 MEC 部署是提供高带宽、低延时、高可靠网络和计算服务的基础。

MEC 服务器部署的选择取决于多种因素，包括可扩展性、物理部署约束、性能指标（如延时）等。此外，部署 MEC 服务器还需要考虑延时、安装成本和服务质量（QoS, quality of service）之间的权衡。相对于边缘云放置技术，MEC 服务部署更加关注固网和移动网络融合架构下部署，如 4G、5G、Wi-Fi，以及多个固定接入网络环境下的 MEC 服务部署方案。具体来说，MEC 服务部署主要可以分为：基于 LTE 网络、5G 网络、以及基于网络控制器的部署方案。

2.2 基于 LTE 网络的部署

在 LTE 网络中，MEC 服务器有两种形态：1) 作为基站的增强功能，通过软件升级或者新增版卡，与基站集成的内置方式。或作为独立设备，将 MEC 服务器部署在无线（RAN）侧；2) 作为独立设备，将 MEC 服务器部署在核心网（CN）侧 [1]。

方案 1：MEC 服务器部署在 RAN 侧。 MEC 服务器部署在多个 eNB 的汇聚节点之后（比较常见的部署方式，见图 4-2-1），或者单个 eNB 之后（针对热点区域，例如校园、大型购物中心等，见图 4-2-2）。这种架构方案的优势在于更方便地通过监听、解析 S1 接口的信令来获取基站侧无线相关信息，但计费 and 合法监听等安全问题需要进一步解决。

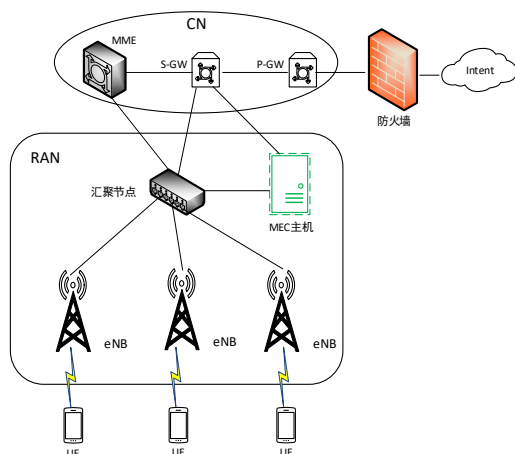


图 4-2-1 MEC 部署在基站汇聚节点后

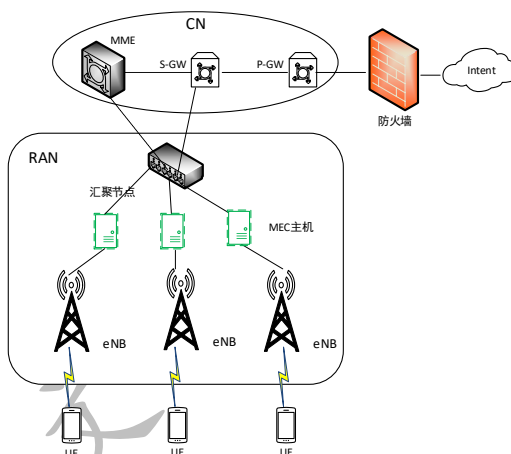


图 4-2-2 MEC 部署在单个基站后

方案 2：MEC 部署在 CN 侧。 MEC 服务器部署在核心网 P-GW 之后（或与 P-GW 集成在一起），可以解决方案 1 中的计费和网络安全等问题。存在的问题是 MEC 服务器的位置离用户较远，时延较大，且占用核心网资源。具体有 2 种实现方案：

(1) 如图 4-2-3，不改变现有 EPC 架构，MEC 服务器与 P-GW 部署在一起。UE 发起的数据业务经过 eNB、HubNode、S-GW、P-GW+MEC 服务器，然后到公网 Internet。(2) 图 4-2-4，改变现有 EPC 架构，MEC 服务器与 D-GW 部署在一起，原 P-GW 拆分为 P1-GW 和 P2-GW（即 DGW），其中 P1-GW 驻留在原位置，D-GW 下移（可以到 RAN 侧，也可以到 CN 边缘）。D-GW 具备计费、监听、鉴权等功能。MEC 服务器可以与 D-GW 集成在一起，也可以作为单独网元部署在 D-GW 之后。P1-GW 与 D-GW 之间为私有接口，需同一厂家设备。

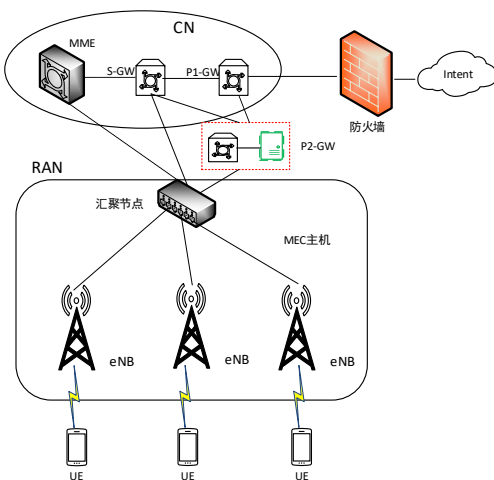
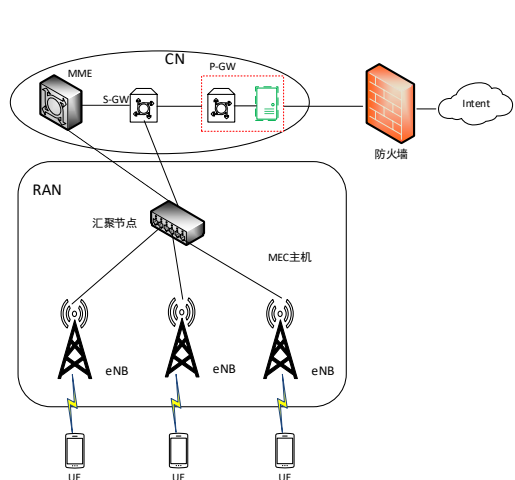


图 4-2-3 MEC 与 P-GW 部署在一起

图 4-2-4 MEC 与 D-GW 部署在一起

2.3 基于 5G 网络的部署

5G 网络的基础设施平台将主要由采用通用架构的数据中心（data center, DC）组成，主要包括中心级、汇聚级、边缘级和接入级，如图 4-2-5 所示，其各自的功能划分大致如下^[2]。

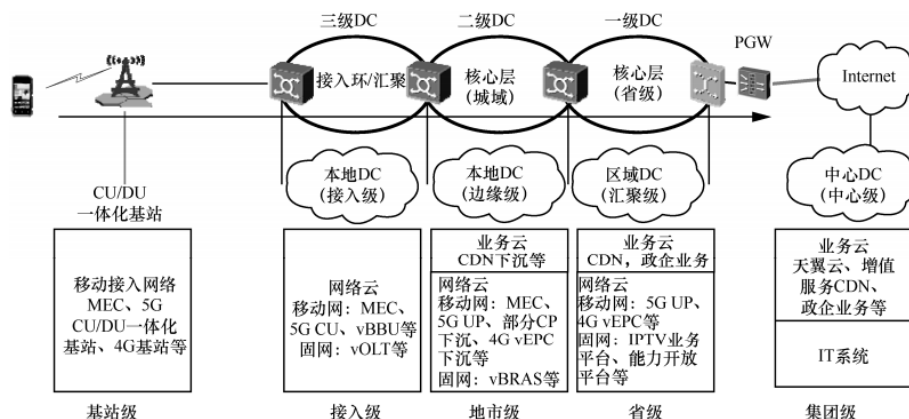


图 4-2-5 5G 网络组网及 MEC 部署策略

(1) 中心级

主要包含 IT 系统和业务云，其中 IT 系统以控制、管理、调度职能为核心，例如网络功能管理编排、广域数据中心互联和 BOSS 等，实现网络总体的监控和维护。除此之外，运营商自有的云业务、增值服务、CDN、集团类政企业务等均部署在中心级 DC 的业务云平台。

(2) 汇聚级

主要包括 5G 网络的控制面功能，例如接入管理、移动性管理、会话管理、策略控制等，主要部署在省级 DC。同时原有 4G 网络的虚拟化核心网、固网的 IPTV 业务平台以及能力开放平台等可以共 DC 部署。除此之外，考虑到 CDN 下沉以及省级公司特有政企业务的需求，省级业务云也可以同时部署在该数据中心。

(3) 边缘级

部署在地市级，主要负责数据面网关功能（包括 5G 用户面功能以及 4G vEPC 的下沉 PGW 用户面功能 PGW-D）。除此之外，MEC、5G 部分控制面功能以及固网 vBRAS 也可以部署在本地 DC。更进一步，为了提升宽带用户的业务体验，固网部分 CDN 资源也可以部署在本地 DC 的业务云里。

(4) 接入级

对于本地接入级 DC，则重点面向接入网络，主要包括 5G 接入 CU、4G 虚拟化 BBU（池）、MEC 以及固网 vOLT 等功能。其中 5G 接入 CU 也可以与其分布式单元（DU）合设，直接以一体化基站的形式出现，针对超低时延的业务需求将 MEC 功能部署在 CU 甚至 CU/DU 一体化基站上。

根据网络传输链路的典型时延值估算，对于 eMBB 场景，MEC 的部署位置不应高于地市级。考虑到 5G 网络用户面功能 UPF 极有可能下沉至地市级（控制面依然在省级），此时 MEC 可以和 5G 下沉的 UPF 合设，满足 5G 增强移动宽带场景对于业务 10 ms 级的时延要求。然而对于超低时延高可靠场景 1ms 的极低时延要求，由于空口传输已经消耗 0.5 ms，此时已经没有给回传留下任何时间。可以理解为，针对 1ms 的极端低时延要求，直接将 MEC 功能部署在 5G 接入 CU 或者 CU/DU 一体

化的基站上，将传统的多跳的网络转化为一跳网络，完全消除传输引入的时延。同时，考虑到业务应用的处理时延，1ms 的极端

时延要求对应的应该是终端用户和 MEC 业务应用间的单向业务，见表 4-2-1。

表 4-2-1 5G 网络典型场景的时延要求

类型	空口单向时延	说明	总体建议
4G	5 ms	基站级: 1~4ms 接入级(三级 DC): 2~3ms 边缘级(二级 DC): 1~4ms 汇聚级(一级 DC): 2~5ms	基于业务需求（时延等），实现 MEC 在不同级别 DC 的部署
5G eMBB	4ms	10ms 级的业务端到端时延，需要降低或者消除传输时延。	MEC 部署在二级 DC（地市），UP 部署于二级 DC（地市），UP/MEC 合设 CP 部署于一级 DC（省级）。
5G uRLLC	0.5ms	1ms 极低时延要求，业务需直接部署接入侧(CU、CU/DU 一体化基站)，消除传输时延。	工业控制根据业务场景和实时性要求，其闭环时延要求在 2~20ms，针对 2~20ms 时延，MEC 可部署的位置主要从基站至接入级 DC；针对 1~2ms 时延要求，MEC 建议部署在 CU 或者 CU/DU 一体化基站上（将多跳转化为一跳）。
5G mMTC	4ms	大规模 MTC 连接场景下，MEC 主要作用体现在通过将 MTC 终端高能耗计算任务卸载到 MEC 平台上，降低终端成本、能耗和延长待机时间。	mMTC 业务主要是利用 MEC 平台的计算、存储能力实现 MTC 终端数据与信令的汇聚及处理，降低网络负荷。因此，MEC 可部署的范围从基站到省级数据中心，甚至可以将其他功能部署在 MTC 终端簇头节点，实现 MTC 终端数据/信令的汇聚处理。

上述仅仅是从时延的角度进行初步分析，当 MEC 应用在企业园区、校园等场景时，考虑到其业务应用服务的覆盖范围以及业务应用数据本地化的需求（出于数据安全性考虑），此时 MEC 则可根据需求部署在该覆盖范围基站的汇聚点，以汇聚网关的形式出现。因此，5G MEC 总的部署策略是应根据业务应用的时延、服务覆盖范围等要求，同时结合网络设施的 DC 化改造趋势，将所需的 MEC 业务应用以及服务部署在相应层级的数据中心。

2.4 基于融合网络的部署

融合网络部署存在两种：（1）4G/5G 无线网络共存环境下，即 eNB 和 gNB 公用基站附近 MEC 服务器；（2）固移融合部署，即支持固定网络和移动网络融合，支持多连接，为业务提供“一致性的业务体验”的融合核心网。

（1）4G/5G 共 MEC 服务器

MEC 联合部署架构目的在于能够同时支持 UE 终端在 4G 和 5G 网络环境下的 MEC 业务服务，如下图 4-2-6^[3]。

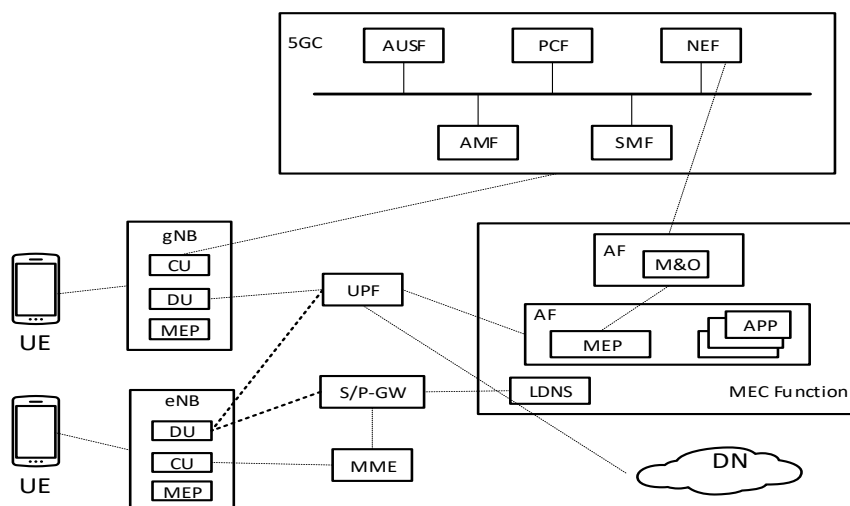


图 4-2-6 MEC 4G/5G 共部署架构图

(2) 基于 MEC 的 5G 融合核心网架构

5G 网络采用分布式网关架构，支持用户面网关 UPF 分布式灵活部署（如 UPF 下沉）以及多个 UPF 级联，基于 5G 网络架构，如下图 4-2-7 所示，基于 MEC 的 5G 融合核心网架构。根据网元部署位置将整个网络分为 3 个层次，用户侧、边缘侧和中心侧。用户侧包括网络的用户终端设备，例如移动网络中的 UE、固定网络中的终端（如用户终端设备（CPE, Customer Premise Equipment）、家庭网关（RG, Residential Gateway Group））等；边缘侧包括靠近用户部署的网络功能，部署位置不高于城域网，主要包括负载接入和汇聚的网络功能，如移动接入网、边缘 UPF、非 3GPP 接入网络、固定网络等；中心侧主要包括移动核心网、固定核心网中的 AAA 和 BPCF 网元，部署位置通常是城域网之上的区域级机房或数据中心。基于 MEC 的 5G 融合核心网架构的基本原则是作为用户面融合点的网关必须部署在网络边缘侧^[4]。

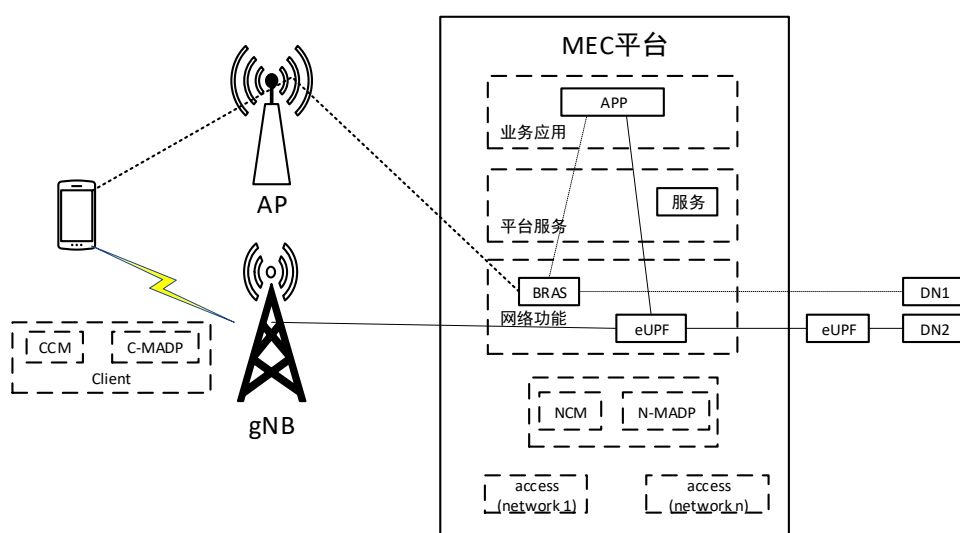


图 4-2-6 基于 MEC 的 5G 融合核心网架构

2.5 参考文献

- [1] 李福昌, 李一喆, 唐雄燕, 等. MEC 关键解决方案与应用思考 [J]. 邮电设计技术, 2016 (11): 81-86.
- [2] 张建敏, 谢伟良, 杨峰义, 等. 5G MEC 融合架构及部署策略[J]. 电信科学, 2018, 34(4): 109-117.
- [3] 陈昕, 温向明, 王鲁晗, et al. 5G 中多接入边缘计算的联合部署架构设计[J]. 北京邮电大学学报, 2018, 41(05):90-95+101.
- [4] 张建敏, 杨峰义, 武洲云等. 多接入边缘计算 (MEC) 及关键技术[M]. 北京: 人民邮电出版社, 2019: 241-242

3 本地分流

3.1 简介

本地分流 (Traffic Offloading) 也叫流量卸载, 基本思想是终端侧或设备侧产生的数据流, 通过某些分流规则和策略, 将部分数据或者大部分数据分流或卸载到 MEC 服务侧, 实现业务数据流本地产生/本地终结、服务本地化。解决政企、校园、工厂、园区、景区、体育馆、博物馆、商场、办公楼以及其他部分垂直行业场景所要求的低时延、高带宽的本地连接以及业务的本地化需求。从而实现了业务应用在无线或有线网络中的本地化、近距离部署、低时延、高带宽传输能力。

3.2 关键技术

本地分流主要考虑的是无线 (4G、5G) 接入网络下的数据分流, 包括家庭/企业网络, 以及 SNS、IMS 业务场景, 主流卸载技术方案包括本地 IP 接入 (LIPA)、选择性 IP 数据流分流 (SIPTO)、IP 移动性 (IFOM) 三类主流技术, 更多的是在核心网络和无线接入环境下的数据 (上下行, 及控制面) 分流。结合 MEC 技术完成本地技术分流, 当前主要的方案是将核心网部分数据面或控制面网元下沉到 MEC 主机系统环境中实现的本地数据分流^[1]。

根据不同组网条件, 可分为以下几类基于 MEC 的分流技术解决方案:

●面向 4G 分流技术的 TOF+方案和 CUPS 方案;

(1) TOF+方案

通过 SGW 下沉到 MEC, LBO (Local breakout) 分流实现。该方案对运营商现网的影响最小, SGW 可以随 MEC 一起部署, 通过 SGW 增强实现 LBO (local break out) 功能。

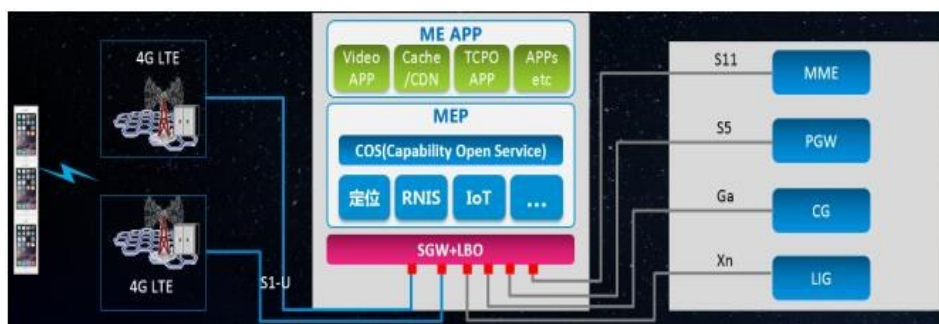


图 4-3-1 SGW 下沉+LBO 分流

如上图所示，将 SGW-U 与 LBO 功能合设（SGW-U 功能增强），支持对上行数据包进行 IP 地址检测，从而实现数据分流和本地卸载。其中 SGW-U 的功能包括：识别上行数据包的 IP 地址，将符合本地分流的数据包传递给 LBO 功能；将从 LBO 过来的下行数据与从核心 PGW-U 过来的下行数据进行整合。实现本地分流流量的计费功能。实现本地分流流量的合法监听功能。LBO 的功能包括：将 SGW-U 分流出的数据，卸载到本地数据网络中，以及将从本地过来的下行数据流封装到对应承载上，以 GTP-U 隧道包传递给 SGW-U。

(2) CUPS 方案

CUPS 方案需要采用 CU 分离方案部署，在核心机房部署 GW-C，在边缘机房部署 GW-U，通过 GW-U 上的分流功能实现 MEC 分流。

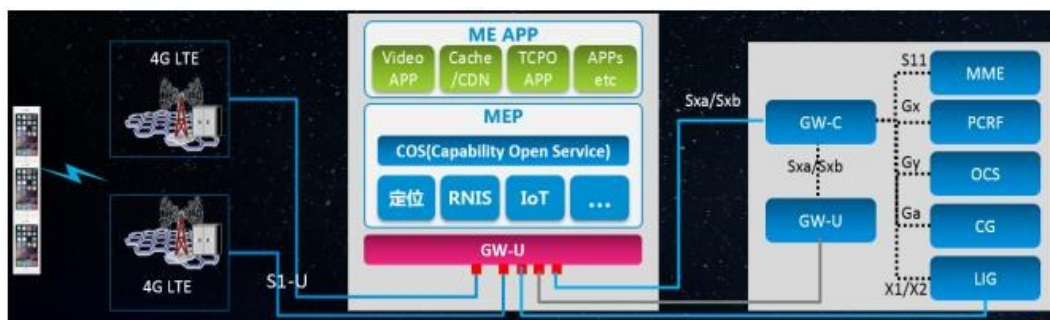


图 4-3-2 CUPS 分流

如图所示，对于 CUPS 分流方案，CU 分离后网关（GW-C+GW-U）的网元功能和外部业务接口无变化，无需改造周边网元（UE、RAN 等），可以与现有网络各元正常对接。GW-C 统一出信令接口连周边设备，简化网络部署。

● 面向 5G 分流技术的 LADN 方案、UL-CL 方案和 Multi-homing 方案。

基于 5G 核心网的 C/U 分离式架构，用户面功能 UPF（user plane function）需要下沉到网络边缘部署，以减少传输时延，实现数据流量的本地分流。控制面功能网元如 SMF（session management function）则在中心 DC 集中部署，便于统一控制部署在 MEC 的 UPF，统一配置和下发分流策略。

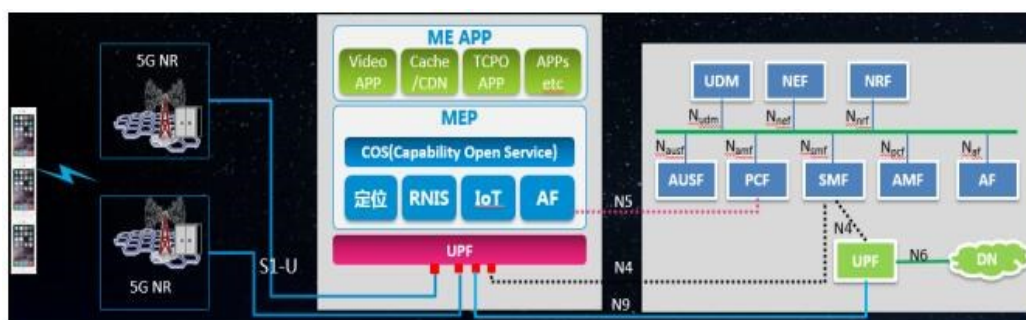


图 4-3-3 UPF 分流

5G 用户面 UPF 下沉部署，实现本地分流方案上图所示。本地 MEC AF 将 UPF 分流规则通过 N5/N33 接口告知 PCF，PCF 将分流策略配置给 SMF，SMF 对所有流量进行集中调度，可采用 LADN（Local Area Data Network）、UL-CL（Uplink Classifier）分流或 Multi-Homing 分流等方案实现边缘 UPF 的分流选择，并将需要分流的本地流量通过本地边缘

UPF 卸载。对于非本地流量则通过本地 UPF 发送到中心 UPF 处理。这样可避免所有流量都迂回中心网络，减轻骨干网传输的压力和建网成本，提升网内分组数据的承载效率与用户业务体验。

(1) LADN

LADN 是和区域服务或应用相关联的 DN 设计，当用户使用该应用时，是通过 LADN 进行访问。当用户位置不在 LADN 的服务区内时，不能接入 LADN，即，通过 LADN PDU 会话接入 DN 只在特定的 LADN 服务区有效。LADN 服务区用一组 TA 标识。支持 LADN 是 5G 支持边缘计算的一种会话管理机制。使用 LADN 用于边缘计算流量分流时，通常 LADN 和单一边缘计算平台的服务区域是一一对应的。



图 4-3-4 LADN 分流

(2) UL-CL 分流

当 PDU 会话类型为 IPv4 或 IPv6 或 IPv4v6 或 Ethernet 时，SMF 可能决定在 PDU 会话的数据路径中插入一个“UL CL” (Uplink classifier)。支持 UL CL 功能的 UPF 通过匹配 SMF 提供的流过滤器将某些流量进行分流。

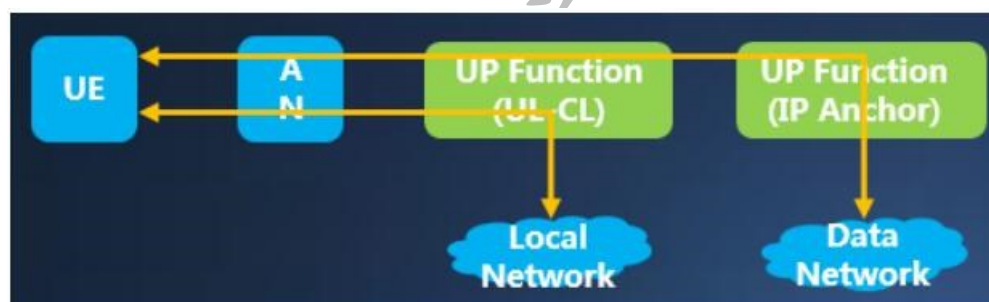


图 4-3-5 UL-CL 分流

(3) Multi-Homing 分流

一个 PDU 会话可能关联多个 IPv6 前缀，这就是 multi-homing PDU 会话。multi-homing PDU 会话，提供通过多个 PDU 会话锚点到 DN 的接入。到不同 PDU 会话锚点的不同的用户面路径，在“common” UPF 形成分支，公共的 UPF 被称为支持“Branching Point”功能的 UPF。Branching Point 转发上行流量到不同 PDU 会话锚点，并聚合发送到 UE 的下行流量，即，聚合从不同 PDU 会话锚点发送到 UE 的流。

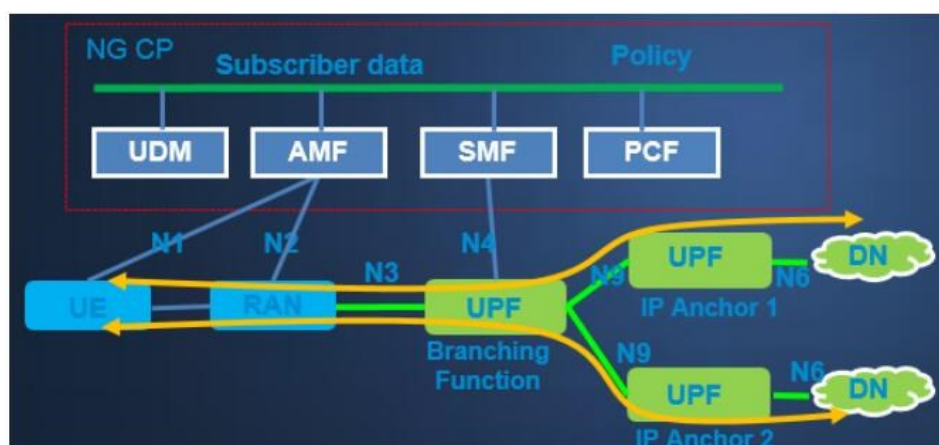


图 4-3-6 Multi-Homing 分流

MEC 应用场景主要是面向 5G 多种业务，如 IoT、车联网、医院、公共安全等，这类场景对业务都具有特殊要求。以及多种多种接入网（如 Wi-Fi、固网有线、ZigBee、NB-IoT 等）场景下的应用。根据业务场景，典型的本地业务分流方案，如基于安防项目的 MEC 视频分流、基于网络的视频 CDN 业务 MEC 数据分流、基于车联网应用业务 MEC 数据分流、基于工业传感设备 MEC 数据分流等

3.3 研究进展

当前本地分流技术的研究还是比较薄弱，且没有形成标准化解决方案。如何根据业务需求和网络接入技术，制定高效的数据流策略和规则是 MEC 本地分流技术首先需要解决的技术问题。其次，在 MEC 本地分流场景下，如何进行本地数据流/内容的计费、合法监听、安全管理、以及差异化的控制，实现运营商、服务提供商、开发者等产业链参与者们的双赢，也是 MEC 商业化部署的需要解决的关键问题。MEC 计算卸载研究进展，如下表所示描述：

刊文	内容概述
3GPP R10、R11 [1][3]	LIPA/SIPTO 技术，实现通过家庭/企业基站（HeNB）进行数据分流、内部往来用户间的直接通信以及宏网络中特定 IP 数据流的直接分流，从而缓解核心网的传输负荷以及投资成本。
Offloading with IFOM: The uplink case [4]	基于 IP 流动性（IFOM）的上行链路卸载。IFOM 允许 LTE 移动用户设备（UE）维持两个并发数据流，一个通过 LTE 而另一个通过 WiFi 接入技术，由于 IEEE 802.11 DCF 的固有公平性设计而限制上行链路。在本文中，提出了两种上行链路卸载算法，以提高 UE 的能效，并在同时使用 IFOM 提供的接入技术的同时增加卸载数据量。
Virtualised EPC for on-demand mobile traffic offloading in 5G environments [5]	介绍了一种基于 NFV 的流量卸载框架体系结构，该体系结构使用 vEPC，旨在当传统 EPC 网络容量达到卸载阈值时实现按需流量卸载。在提出的框架体系结构中，同时提出了三种表示不同卸载方式和资源利用率的 vEPC 模式。最后还指定了利用流量监控方法，按需实例化 vEPC。
移动边缘计算技术及其本地分流方案 [6]	

	首先介绍 MEC 技术，细化并给出 MEC 平台框图。此外，针对基于 MEC 平台的本地分流功能，给出了详细的技术方案，并与 3GPP 本地分流方案 LIPA/SIPTO 进行对比分析。
刊文	内容概述
基于 MEC 的 LTE 本地分流技术 ^[7]	本文提出了基于 MEC 的 LTE 本地分流技术方案。同时，通过 LTE 主要信令流程的解析，深入分析了该技术方案的可性。相比于 3GPP 现有本地分流方案 LIPA/SIPTO，基于 MEC 的 LTE，本地分流方案对终端与网络是透明的，更适合 LTE 现网本地分流业务的部署。终端无需支持多个 APN 连接，从而为用户提供一种“虚拟的 LTE 局域网”体验。
刊文	内容概述
Smart Traffic Offloading with Mobile Edge Computing for Disaster-Resilient Communication Networks ^[8]	灾难恢复通信网络中，基于移动边缘计算实现的智能流量卸载技术。发生大规模自然灾害时，灾区的 eNB 节点可能被破坏，且存在核心网（CN）数据流大量拥塞情况。在本文中，利用在车载 eNB（VeNB）上一种智能流量卸载机制（STOM），以提高 CN 灾害恢复通信系统的吞吐量和减少网络延时。
刊文	内容概述
Traffic offload improved method for 4G/5G mobile network operator ^[9]	4G / 5G 运营商网络流量卸载的一种改进方法,本文提出了三种流量卸载方案，并对其进行了建模，数据显示有效流量卸载达到 55%。
刊文	内容概述
A study on M-CORD based architecture in traffic offloading for 5G-enabled multiaccess edge computing networks ^[10]	基于 M-CORD 架构支持 5G MEC 流量卸载的研究。本文提出一种新的架构，使计算更接近 5G 边缘网络中的用户。
刊文	内容概述
Traffic Offloading Algorithm Using Social Context in MEC Environment ^[11]	一种基于 MEC 服务网络环境社交内容的流量卸载算法，LIPA 和 SIPTO 卸载技术可以满足 QoS 移动业务要求。由于 SNS 大量业务产生大量流量，基于社交环境的流量卸载是有益的。本方案利用社交应用程序使用率，和流行程度、以及根据有效数据速率大小，找到最优卸载权重因子，以及最大化小区用户服务质量。最后，通过应用选择概率和最优卸载权重因子确定卸载比。通过性能分析，相当对传统算法可以有效提升 46%卸载。
刊文	内容概述
Mobile Edge Computing Solutions for IMS Traffic Offloading in 4G/5G Mobile Networks ^[12]	为了降低核心网端到端的时延和业务负载，本文提出了一种 IMS-ALG（IMS 应用层网关）下层到 MEC 方案，MEC 设备修改 ip/port 地址以更改后续流媒体传输路径。本文改进了 MEC 设备上的 ALG 机制，在没有 MGW 转发情况下，在本地网络中传输媒体流。通过去除冗余传输路径，可以提高媒体传输性能。

3.4 参考文献

- [1] “灵活高效的 MEC 分流方案，使能行业客户定制虚拟移动专网”，中兴通讯 MEC 分流方案[EB/OL],
<https://www.icloudnews.net/a/20190819/22683.html>.
- [2] 3GPP. Local IP access and selected IP traffic offload (LIPA-SIPTO) (release 10): TR 23.829[S]. 2011.
- [3] 3GPP. LIPA mobility and SIPTO at the local network (release 11): TR 23.859[S]. 2011.

-
- [4] V. Miliotis, L. Alonso, and C. Verikoukis, "Offloading with ifom: The uplink case," in *2014 IEEE Global Communications Conference*, Dec 2014, pp. 2661–2666
- [5] Seil Jeon ; Daniel Corujo ; Rui L. Aguiar. "Virtualised EPC for on-demand mobile traffic offloading in 5G environments" ,IEEE .TCSET, 2015
- [6] 张建敏, 谢伟良, 杨峰义等, 移动边缘计算技术及其本地分流方案[J], 电信科学 2016 年第 7 期。
- [7] 张建敏, 谢伟良, 杨峰义等, 基于 MEC 的 LTE 本地分流技术[J], 电信科学 2017 年第 6 期。
- [8] Wen-Pin Che,Ang-Hsun,Chung-Hsien, "Smart Traffic Offloading with Mobile Edge Computing for Disaster-Resilient Communication Networks" , Journal of Network and Systems Management. April 2019, Volume 27, Issue 2, pp 463–488
- [9] Roman Odarchenko ; Anastasiia Abakumova ; Oleh Polihenko. "Traffic offload improved method for 4G/5G mobile network operator" , IEEE .TCSET, 2018.
- [10] Kathiravan Srinivasan ; Nitesh Kumar Agrawal, "A study on M-CORD based architecture in traffic offloading for 5G-enabled multiaccess edge computing networks" , IEEE.ICASI. 2018.
- [11] Hye-Rim Cheon et al., "Traffic Offloading Algorithm Using Social Context in MEC Environment", Journal of KICIS, vol. 42, no. 02, Feb. 2017.
- [12] Whai-En Chen,Li-Yao Tseng, "Mobile Edge Computing Solutions for IMS Traffic Offloading in 4G/5G Mobile Networks" , NCS 2017,P12 – 17.

4 缓存加速

实际网络中, 存在大量用户频繁访问相同的业务数据 (如热点视频、新闻等), 如果业务数据流需要经过终端、基站、回传网络、核心网、以及业务服务器等长距离的传输, 必将导致时延增加, 用户体验下降, 同时带宽资源浪费、造成链路拥塞等问题。因此, 根据业务需求、用户习惯、以及业务类型等将部分内容缓存在本地供用户访问, 可以有效的降低业务时延、节省传输带宽, 提升用户体验。MEC 技术是将通用服务器部署在接入侧边缘, 地理位置上更将接近用户, 因此, 研究基于 MEC 技术的缓存技术具有重要意义和商业前景。本章节首先梳理现有的传统缓存加速策略, 然后在介绍 MEC 应用环境下的缓存加速策略研究进展。

4.1 CDN 简介

内容分发网络 (Content Distribution Networks, CDN) 是一种基于互联网的缓存网络, 通过在网络边缘部署缓存服务器来降低远程站点的数据下载延时, 加速内容交付。缓存加速策略一直都是学术界研究的热点^[1]。

以内容为网络通信主体的信息中心网络 (ICN, information-centric networking), 专注于内容的分发和检索, 而不是特定端点间的信息传递。按照缓存指标可以将传统缓存策略分为五大类, 如下图 4-4-1 所示, 这些缓存策略研究了传输路径上的缓存方法, 在网络层中增加了命名和缓存机制, 并在路由器上执行, 从而减少网络中所需的流量并且降低时延。因此,

及时、合理地将内容放置，是提升缓存效率的有效途径。基于移动网络边缘的缓存加速侧包括：基于用户偏好的缓存加速策略、基于学习的缓存加速策略、非协作式缓存加速策略、协作式缓存加速策略[2]。

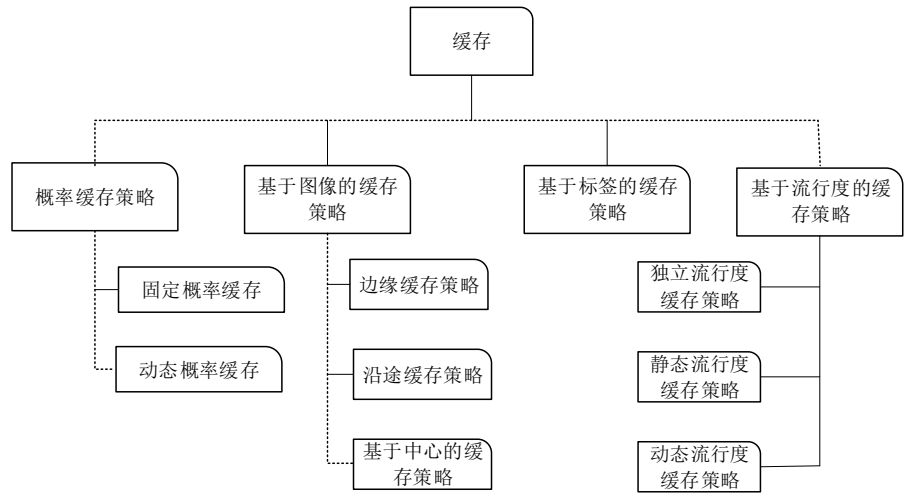


图 4-4-1 传统缓存加速策略概览

目前，运营商网络中的 CDN 边缘节点多部署于数据网络的核心层和业务控制层，大部分为固网 IPTV 业务专用，融合 CDN 尚处于探索阶段；而互联网 CDN 则多部署于 Internet 骨干网，无法真正靠近移动用户。随着无线网络的发展，特别是 5G 技术对大带宽和实时性需求，CDN 技术引入 MEC 服务器部署成为另一种新型的部署策略，结合 MEC 技术和 CDN 技术俨然成为新的缓存加速策略。

4.2 研究进展

针对缓存加速的研究有很多，当前研究方向大部分是结合了 5G、NFV、MEC、SDN。本文罗列出部分 MEC CDN 研究进展，如下表所示描述。

刊文	内容概述
面向视频流的 MEC 缓存转码联合优化研究[3]	本文从 MEC 和 ABR 的背景和概述出发，对目前面向视频流的缓存转码资源联合优化方案进行了介绍和分析，并主要从缓存策略、转码策略和优化方式等方面对已有方案进行了对比。在对以上方案分析对比的基础上，研究了面向视频流的 MEC 资源优化问题目前面临的挑战和研究难点，如与时延优化的均衡问题、能量优化问题和用户行为分析以及分布式 MEC 协作的问题等，在网络整体优化方面具有重要意义。
刊文	内容概述
移动边缘计算中数据缓存和计算迁移的智能优化技术[4]	针对移动边缘计算（MEC），设计了计算迁移和数据缓存的联合优化模型，并基于改进的遗传算法（GA）对该模型的时延优化特性进行了优化，有效提高了缓存空间的使用效率，性能方面也有较大程度的提升。
刊文	内容概述
基于 5G 边缘网络的视频 CDN 业务分流方案探讨[5]	在分析 5G 网络架构和 CDN 现状的基础上，对移动 CDN 在 5G 边缘网络中的部署进行了深入研究和讨论，对本地业务分流提出了解决方案（基于 PCC 的策略配置，基于 MEC 本地配置的分流，基于第三方应用的自主分流）并进行了对比分析，对不同场景下的方案选择提出了建议。
刊文	内容概述

面向 5G MEC 边缘云的 CDN 下沉方案 ^[6]	本文重点介绍了基于边缘云的 vCDN 的系统架构实现方案，设计了相关的业务流程，开展了验证试验，最后给出了建设部署方案。测试结果表明，将应用程序部署到网络边缘不仅能够降低回传带宽压力和延迟，而且优化了用户体验，同时帮助业务提供方降低了成本。
刊文	内容概述
MEC Proxy for Efficient Cache and Reliable Multi-CDN Video Distribution ^[7]	基于 MEC 代理实现的高效缓存和可靠的 Muti-CDN 视频分发技术。本文提供了一种新的 MEC 代理来扩展和部署缓存架构，以实现高效可靠的内容分发。第一，代理在网络边缘缓存内容，以减少 OTT 服务提供商 CDN 的资本支出 (CAPEX)。为此，代理能够识别重复请求。其次，代理可以防止已识别或预测的 CDN 故障。因此，代理将下载会话然后切换到备用 CDN，以确保 QoE 速率，从而启用基于实时连接统计的 CDN 动态选择。通过在密集的客户单元中通过 HTTP 动态自适应流 (MPEG-DASH) 来评估所提出的解决方案。
刊文	内容概述
CDN Slicing over a Multi-Domain Edge Cloud ^[8]	多域边缘云上的 CDN 切片方案。基于 5G 网络技术中的 NFV 和 MEC 技术，引入 CDN 服务技术，且符合 NFV 和 MEC 技术标准。为实现服务质量 (QoE) 优化，推导出 QoE 服务工作量变更情况下的服务模型。再加上多级服务监控，设计出 CDN 模型的资源分片和弹性管理机制。能够实现自动计算资源并伸缩和动态转码，以适应当前的视频比特率网络状态。最后，通过实验验证了方案的有效性。
刊文	内容概述
Integrated ICN and CDN Slice as a Service ^[9]	本文利用 NFV 和 MEC 技术，提出了一种将 ICN (信息中心网络) 与 CDN (内容交付网络) 集成在一起的系统，以提供有效的内容交付服务。将动态 CDN 切片概念与基于 NDN (命名数据网络) 的 ICN 切片概念相结合以避免核心网络拥塞。部署动态 CDN 切片可以根据内容特点和潜在观众的地理分布实现在最佳位置缓存内容。虚拟缓存服务器与支持虚拟转换器一起放置在属于多个管理域的云上，形成一个 CDN 分片。ICN 分片则用于内容的区域分布，利用基于名称的访问和网络内容缓存的自主性，使得能够从附近的网络节点传递内容，从而避免内容的重复传输，并获得更短的响应时间。我们的实验表明，ICN / CDN 的集成，在几乎所有方面 (包括服务可伸缩性、可靠性和服务质量) 都优于传统 CDN。
刊文	内容概述
Congestion Avoidance and Load Balancing in Content Placement and Request Redirection for Mobile CDN ^[10]	移动 CDN 中的内容放置和请求重定向的拥塞避免和负载均衡。在一个移动通信系统中，基站 (BSs) 配备 CDN 内容存储，是比较广泛机制。本文研究了基于 BSs 的移动 CDN 的内容放置和请求重定向的优化问题。具体地说，每个 BS 维护一个传输队列，用于应答从其他 BS 发出的请求。通过保证网络的稳定性，同时考虑网络拥塞和 BSs 负载均衡。在网络稳定性约束下，采用随机优化模型来最小化平均传输成本。利用 Lyapunov 优化技术，我们将长期问题转化为一组在短时间求解的线性规范，并且开发了一种在线算法，可以有效地确定内容放置和请求重定向，而无需判定随机网络状态信息。通过理论分析，给出了该算法的最优性和网络稳定性方面的性能。评估证实，该解决方案可以有效较低的传输成本，同时避免拥塞和负载均衡。
刊文	内容概述
Optimal VNFs Placement in CDN Slicing Over Multi-Cloud Environment ^[11]	本文介绍了一种内容交付网络即服务 (CDNaaS) 平台，该平台允许动态部署和跨多个管理云域运行的虚拟内容交付网络 (CDN) 片的生命周期管理。CDN 切片由四种虚拟网络功能 (VNF) 类型组成，即虚拟代码转换器，虚拟流媒体，虚拟缓存和 CDN 切片专用的协调器，用于管理所涉及的云域中的切片资源。为了创建高效的 CDN 切片，对每个 VNF 使用足够数量的虚拟资源，并对 VNF 进行优化。本文设计了一种机制，可以为每个 CDN 切片分配一组合适的 VNF，以满足其性能要求，并尽可能地降低所分配的虚拟资源成本。然后，建立了数学模型来评估所提出机制的性能，通过建立数学模型来评估所提出机制的性能，大量仿真实验证明可以有效实现目标设计。
刊文	内容概述

An Edge Content Delivery and Update Framework in Mobile Edge Computing ^[12]	本文提出了在物联网 (IoT) 环境中, 基于移动边缘计算的边缘内容交付框架 (ECD), 以减轻核心网络的负载并提高移动用户的体验质量 (QoE)。基于内容流行度和缓存池排名, 提出了边缘内容交付方案和边缘内容更新方案。
刊文	内容概述
Satellite Support for Enhanced Mobile Broadband Content Delivery in 5G ^[13]	卫星通信作为 5G 回传的使能技术之一, 特别是在 5G 中交付带宽需求增强的移动宽带 (eMBB) 应用数据。本文介绍了一种基于卫星通信的面向 5G 的网络架构, 用于支持新兴的移动视频传输, 并在欧盟 5GPPP 第 2 阶段 SAT5G 项目中进行了研究。同时介绍了两个互补的用例, 包括: (1) 使用卫星链路支持 5G 移动边缘的流行视频内容的离线多播和缓存; (2) DASH (HTTP 上的动态自适应流) 视频段的实时预取。在这两种情况下, 目的是将内容对象卫星通信链路定位在靠近消费者的地方, 以确保 5G 内容应用中的体验质量 (QoE)。

边缘缓存加速, 通过缓存策略、vCDN、视频转码等, 并通过融合现代网络技术如 NFV、SDN、切片、卫星通信技术等, 研究新型 CDN 技术是大趋势, 也是未来学术研究的热点方向。

MEC 服务器与业务系统对接, 获取业务中的热点内容, 包括视频、图片、文档等进行本地缓存。在业务进行过程中, MEC 服务器对基站侧数据进行实时的深度包解析, 如果终端申请的业务内容已在本地缓存中, 则直接将缓存内容定向推送给终端。基于当前新型的 CDN 研究进展基本分析, 总结如何基于 MEC 实现缓存加速方案。

基本思路就是, 通过将缓存加速服务 (Traffic Caching Accelerating Service, TCAS) 部署在 MEC 平台上, 提前缓存终端用户所需的热点内容, 从而实现终端用户所需内容的就近访问, 降低延时。基本思路是 MEC 平台通过对业务数据分组进行解析, 并根据特定的规则对热点内容进行边缘缓存。于此同时, 当用户访问互联网内容时, TCAS 首先会对数据分组进行解析, 对于命中内容, 通过缓存控制机制使数据分组直接以 IP 方式从缓存服务器下载。对于未命中内容, 数据分组由 TCAS 发起从源服务器下载^[1]。TCAS 服务设计需要考虑如下问题:

- (1) 如何最优化部署 TCAS 服务;
- (2) 如何根据差异化业务内容服务, 设计符合业务需求的 TCAS 内容缓存策略、缓存模式, 内容生成算法等;
- (3) 如何高效利用业务缓存资源 (如虚拟存储)、计算资源 (如虚拟 CPU), 带宽资源等;
- (4) 如何降低对 MEC 平台的要求, 如不额外增加传输资源、不增加与互联网的业务链等;
- (5) 如何降低回传网络, 以及业务服务器的压力等;
- (6) 如何充分利用现代网络技术 (如 NFV、SDN、5G), 以及编排技术, 实现 CDN 内容的精准缓存和最优化传输;

4.3 参考文献

- [1] M. Zhang, H. Luo, and H. Zhang, "A Survey of Caching Mechanisms in Information-Centric Networking," IEEE Commun. Surv. Tutor.vol. 17, no. 3, pp. 1473–1499, thirdquarter 2015.
- [2] 张建敏, 杨峰义, 武洲云等. 多接入边缘计算 (MEC) 及关键技术[M]. 北京: 人民邮电出版社, 2019: 132-133.

-
- [3] 李佳, 谢人超, 贾庆民, 等. 面向视频流的 MEC 缓存转码联合优化研究[J]. 电信科学, 2018, 8.
- [4] 汪海霞、赵志峰、张宏纲, 移动边缘计算中数据缓存和计算迁移的智能优化技术[J]. 中兴通讯技术. Apr. 2018, Vol.24 No.2.
- [5] 张伟强, 韦广林, 孙森田, 等. 基于 5G 边缘网络的视频 CDN 业务分流方案探讨 [J] . 邮电设计技术, 2019 (7) : 11-14.
- [6] 吕华章, 王友祥, 唐雄燕. 面向 5G MEC 边缘云的 CDN 下沉方案[J]. 移动通信, 2019,43(1): 20-28.
- [7] Viola, Roberto, et al. "MEC Proxy for Efficient Cache and Reliable Multi-CDN Video Distribution." 2018 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB). IEEE, 2018.
- [8] Taleb, Tarik, et al. "CDN slicing over a multi-domain edge cloud." IEEE Transactions on Mobile Computing (2019).
- [9] Benkacem, Ilias, et al. "Integrated ICN and CDN Slice as a Service." 2018 IEEE Global Communications Conference (GLOBECOM). IEEE, 2018.
- [10] Liu, Jiayi, Qinghai Yang, and Gwendal Simon. "Congestion avoidance and load balancing in content placement and request redirection for mobile CDN." IEEE/ACM Transactions on Networking 26.2 (2018): 851-863.
- [11] Benkacem, Ilias, et al. "Optimal VNFs placement in CDN slicing over multi-cloud environment." IEEE Journal on Selected Areas in Communications 36.3 (2018): 616-627.
- [12] Wang, Shangguang, et al. "ECD: An Edge Content Delivery and Update Framework in Mobile Edge Computing." arXiv preprint arXiv:1805.10783 (2018).
- [13] Wang, Ning, et al. "Satellite support for enhanced mobile broadband content delivery in 5g." 2018 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB). IEEE, 2018.
- [14] 张建敏, 杨峰义, 武洲云等. 多接入边缘计算 (MEC) 及关键技术[M]. 北京: 人民邮电出版社, 2019: 146-152.

5 移动性管理

5.1 简介

移动性管理(Mobility Management), 指的是跟踪移动用户设备 (UE) 并将其与适当的基站 (BS) 相关联, 使得移动系统能够交付数据和服务。这一技术已经广泛应用于传统的 4G、5G 无线网络中, 指的是在不影响语音通话和数据传输的情况下, 移动用户能够自由地从一个地理位置迁移到另一个地理位置, 前后两个位置之间可以存在多个覆盖范围有限的蜂窝基站, 通过各种算法和技术, 能够实现动态移动性管理并保证高速率和低误码率。

早期的 MEC 也称为移动边缘计算（Mobile Edge Computing），就是为了 UE 设备上的移动边缘应用程序提供 IT 服务环境和计算功能，从而可以将业务本地化，在接入网处理用户请求。在移动边缘计算环境下，单个边缘服务环境覆盖范围有限，终端用户（汽车、个人移动终端设备等）频繁移动等，导致边缘服务 QoE 下降，甚至服务中断，无法保障服务的连续性。因此，MEC 移动性管理要求是如何保证用户在移动过程中获得服务的连续性，或是如何使得 MEC 服务能够支持移动性或者保持 MEC 服务的连续性。

5.2 MEC 移动性管理

针对移动性管理，“ETSI GR MEC 018”标准针对端到端的移动性支持有专门的研究和说明^[1]。MEC 场景下的移动性主要包括如下 3 种场景，如下图 4-5-1 所示。

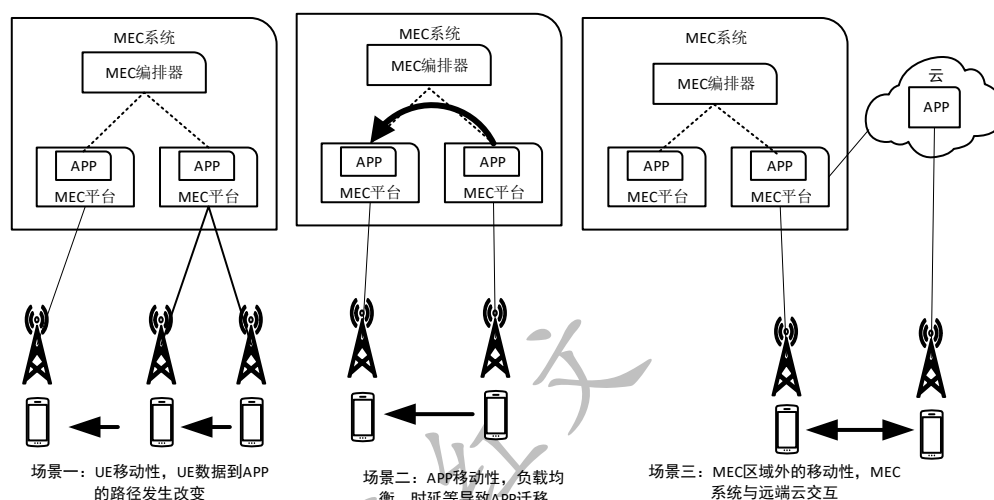


图 4-5-1 MEC 场景下的移动性分类

结合 5G 网络架构，给出如下终端用户可能面临的移动性问题，具体说明可以参考“MEC 关键技术”^[2]。如下图 4-5-2，MEC 场景下移动性示意图。

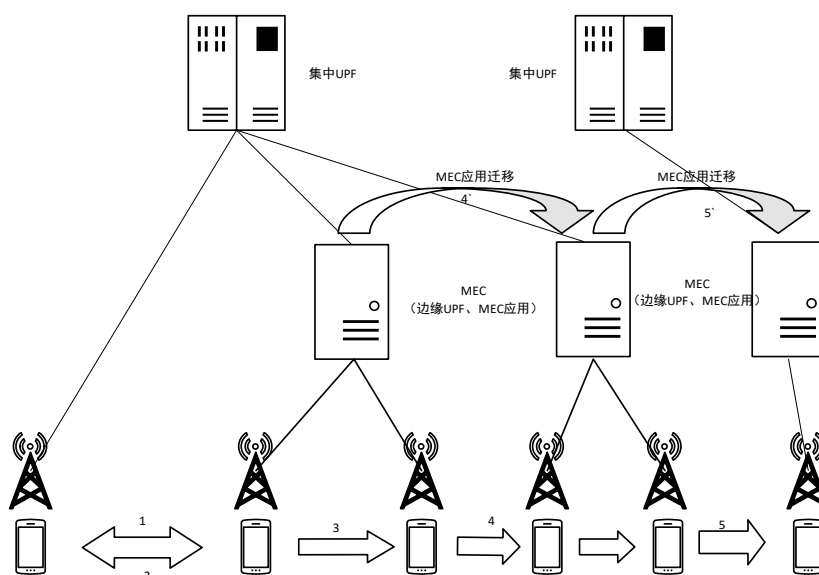


图 4-5-2 MEC 场景下的移动性示意图

5G MEC 场景下的移动性问题可以划分为如下 5 类：

- (1) 同一集中 UPF 下，用户从非 MEC 覆盖区进入 MEC 覆盖区
- (2) 同一集中 UPF 下，用户从 MEC 覆盖区进入非 MEC 覆盖区
- (3) 同一集中 UPF 下，用户在同一 MEC 覆盖范围内跨基站切换
- (4) 同一集中 UPF 下，用户在不同 MEC 覆盖范围内跨基站切换
- (5) 用户在不同的集中 UPF 间切换

在上述问题（3）中，终端用户虽然经历了不同的基站间切换，但仍然在同一个 MEC 服务覆盖范围，此时无需进行 MEC 应用迁移。然而对于问题（4）和（5），终端用户的移动导致其在不同的 MEC 服务覆盖范围下进行切换，除了涉及底层网络中的基站以及边缘 UPF 间的切换，由不同 MEC 应用覆盖范围以及终端用户切换后导致的业务应用访问时延增加等问题，此时需要进行跨 MEC 平台间的业务应用迁移，从而避免 MEC 间切换导致的业务中断，满足业务连续性的需求。以车联网场景为例，用户高速移动以及低延时的要求，给 MEC 业务连续性提出了更大的挑战。

5G MEC 场景下的移动性管理问题归纳为如下技术点：

- (1) 基站间切换，不涉及集中 UPF（用户面功能）重新分配；
- (2) 基站间切换，涉及 UL CL（上行分类器）或 BP 分支点功能的添加与删除；
- (3) 基站间切换，涉及边缘 UPF 间切换；
- (4) MEC 应用迁移（或应用连续性）；
- (5) 基站间切换，涉及集中 UPF 重选。

5.3 应用迁移分类

除了 5G 网络，其他无线网络场景如 4G 场景也存在同样的移动性管理问题。假设每个基站都会连接一个 MEC 主机，那么这些迁移主要概括为：

- (1) 不同基站（4G/5G）小区切换，相同的 MEC 主机；
- (2) 不同基站（4G/5G）小区切换，不同的 MEC 主机；

这些无线场景中为支持 MEC 服务的连续性，都需要解决在无线网络发生迁移或快速移动时 MEC 应用迁移（或者 MEC 服务迁移）问题。相对于无线网络的移动性管理，蜂窝切换主要是通信的切换，这个过程涉及的数据可能较少；而 MEC 移动性管理可能是 MEC 应用（MEC 服务）整体的迁移，涉及到程序、数据、执行状态、和服务状态等从一个 MEC 主机迁移到另一个 MEC 主机上。针对需要迁移的 MEC 业务应用，可以分为以下几种类型^[2]：

- 1) 根据 MEC 业务应用服务属性可以分为专用型业务应用和共享型业务应用；

- 专用型业务应用：此类 MEC 业务应用实例专门服务特定用户，当终端用户移动至其他 MEC 覆盖范围下，该 MEC 业务应用实例需要迁移至目标 MEC 主机上。

● 共享型业务应用：此类 MEC 业务应用实例并非专门服务与特定用户，而是同时服务于多个用户（如多播业务等）或者 MEC 覆盖范围内所有 MEC 用户（如广播业务等）。当终端用户移动至其他 MEC 覆盖范围内时，如果 MEC 业务应用已经在目标 MEC 主机上进行实例化，则无需再进行该 MEC 业务应用的实例化，仅需将该 MEC 业务应用中用户上下文信息传递至目标 MEC 主机即可。同时，考虑到即使终端用户已移出源 MEC 覆盖范围，但还需服务于其他用户，所以无需对其 MEC 业务应用实例进行终结。

2) 根据 MEC 应用是否需要存储用户上下文信息，分为无状态业务应用和有状态业务应用。

● 无状态型业务应用：此类业务应用无需记录或者存储服务状态以及用户相关数据，新的服务会话与之前服务会话的上下文信息无关。

● 有状态型业务应用：此类业务应用可以记录服务会话相关状态信息，该状态信息可以存储在终端用户应用或者 MEC 业务应用实例中，便于在服务会话转化时保证业务连续性。

由此可以看出，相对于虚拟机实时迁移或者镜像实时迁移，MEC 应用迁移业务服务的连续性，虚拟机（或镜像）的迁移更关注虚拟机宕机时间。服务的连续性，考虑的是迁移过程中的整体时间，包括端到端的业务延时和虚拟机（或镜像）迁移时间。另外，还需要考虑业务数据、用户群体的动态性、带宽和无线信号强弱、MEC 主机本身计算能力和资源的变化等等问题。因此，MEC 应用迁移是个非常复杂的过程，这也是目前重点需要研究的方向之一。

5.4 研究进展

目前，针对 MEC 移动性管理的研究，主要包括 MEC 应用迁移（relocation）和 MEC 应用虚拟机（或者镜像）的切换（handover），围绕终端用户在移动性过程中，MEC 应用迁移决策算法、迁移成功率、预测等领域进行研究。如下表所示：

刊文	内容概述
移动边缘计算的移动性管理研究 ^[3]	提出了移动边缘计算（MEC）中基于马尔可夫决策过程（MDP）的虚拟机迁移策略。该策略可以在保证用户体验的前提下，能使得系统成本达到最优。每次用户进入新的 MEC 区域后，都通过该策略决定是否进行迁移。通过与常规迁移策略对比，仿真结果表明本策略总能达到最优情况，有效地降低了系统成本。
刊文	内容概述
移动边缘计算环境下的服务迁移策略设计与实现 ^[4]	针对服务分配中存在的网络延迟高及各个服务器间负载不均衡的问题，提出了一种基于簇划分机制的服务分配算法。与传统启发式算法相比，提高了 MEC 服务器间的负载均衡程度约 20%。其次，为了解决服务迁移中是否进行迁移及迁移目标选择的盲目性引起的迁移开销较大的问题，提出了一个基于强化学习的服务迁移决策方法。首先将服务迁移视为马尔可夫决策过程。在此基础上，结合通信开销、迁移开销、用户移动方向以及服务器可用资源的影响，定义了收益函数。最后通过模拟用户移动数据训练确定迁移决策。仿真结果表明，该方法起到了减少迁移开销的作用。
刊文	内容概述
移动边缘计算缓存优化与用户移动性预测研究 ^[5]	论文基于 MEC 的多层次缓存架构，针对降低用户请求内容的时延，提高缓存效率等优化目标，利用多播技术和基站间的协作机制，提出了一种协作式的多播缓存策略（CMAC, Cooperative Multicast-Aware Caching），并引入了贪婪算法，通过与简单多播缓存（MAC, Multicast-Aware Caching）的性能比较，验证了 CMAC 的有效性，与 MAC 相比平均时延降低了 13%。论文还提出了基于 MEC 的移动性预测框架 MPaS (Mobility Prediction as Service)，采用 LSTM (Long Short-Term Memory) 对用户的移动性进行预

	测,基于公开数据集对 LSTM 进行训练和仿真,比较了 ANN(Artificial Neural Network)和 LSTM 预测的性能,验证 LSTM 对用户移动性预测的有效性。
刊文	内容概述
一种基于差异化服务的 MEC 应用实例智能迁移机制 ^[6]	本文基于差异化服务的 AI 智能迁移机制对 ETSI 原有的机制进行了改进与增强,通过合并可能同时需要发生切换的其他 UE 或者 App 请求,为 AI 迁移定义了 4 种可选的同步模式供系统选择,从而最大限度地实现了特定情况下同步消息数量的减少和切换效率的提升,最终实现了系统对差异化服务的支持。
刊文	内容概述
A Mobility Management Architecture for Seamless Delivery of 5G-IoT Services ^[7]	本文重点介绍跨异构网络的漫游移动用户无线资源的访问。提出了一种统一服务架构,该架构可通过网络切片实现在 5G (新一代核心网) 服务和 4G (演进分组核心网) 服务之间无缝切换。
刊文	内容概述
Exploitation of Mobile Edge Computing in 5G Distributed Mission-Critical Push-to-Talk Service Deployment ^[8]	本文提出了一种非独立的基于 5G ETSI MEC 的架构,用于关键任务一键通 (MCPTT) 服务。提出了一种分层的分布式 MCPTT 体系结构,该体系结构在边缘分配用户平面,同时保持控制平面 (CP) 用于同步和辅助目的。由于功能服务器靠近最终用户,因此 MEC 架构可实现低延迟服务的部署,从而使用户设备免于繁重的工作量,水平扩展性的好处是可以在特定位置动态分配网络资源。所提出的架构在隔离的 E-UTRAN 操作情况下也是有益的,在这种情况下,在回程连接丢失的情况下,可以将本地 MCPTT 服务直接部署在隔离的 eNodeB 组中。
刊文	内容概述
Dynamic Service Migration in Mobile Edge Computing Based on Markov Decision Process ^[9]	基于马尔可夫决策过程的动态服务迁移,服务迁移问题公式化为马尔可夫决策过程 (MDP),公式描述了通用成本模型,并提供了数学框架来设计最佳服务迁移策略。为了克服与计算最佳策略相关的复杂性,通过用户和服务位置之间的距离来近似基础状态空间。证明了生成的 MDP 对于统一的一维用户移动性是精确的,同时为均匀的二维迁移率提供了一个近似值,并具有恒定的加性误差。还提出了一种用于计算最优解的新算法和一种数值技术,它比基于标准值或策略迭代的传统方法要快得多。
刊文	内容概述
Application Synchronization Among Multiple MEC Servers in Connected Vehicle Scenarios ^[10]	本文研究了 MEC 环境中诸如视频实时流服务之类的应用程序级延迟。在实验室内构建的 MEC 环境中进行的实验表明,应用程序级的延迟减少了 15% 至 30%,通过将远程服务功能重新定位在靠近用户的边缘,可以改善应用程序级的延迟。
刊文	内容概述
Efficient Dynamic Service Maintenance for Edge Services ^[11]	应用 Skyline 图模型并采用有向无环图理论来存储和更新移动边缘服务。具体而言, Skyline Graph (SG) 算法旨在解决移动边缘服务的插入,删除,更新和搜索,以实现边缘服务的高效维护。在真实世界的 Web 服务和模拟数据集上都进行了全面的实验,以评估我们方法的有效性和效率。

通过 MEC 场景下的移动性管理,从而保证 MEC 业务的连续性是 MEC 技术需求的亮点之一,这也是区别其他边缘计算的重要特性。MEC 移动性管理,不仅仅考虑 5G 移动性管理,还得考虑 MEC 业务在终端和主机侧的移动性管理。特别是在 5G 环境的低延时 MEC 业务,如何有效降低 MEC 应用迁移导致的延时增加,保证 MEC 服务的容错迁移,MEC 研究还有诸如可预测性、迁移成功率、QoS、计费等诸多问题需要解决。

5.5 参考文献

- [1] “ETSI GR MEC
018,” https://www.etsi.org/deliver/etsi_gr/MEC/001_099/018/01.01.01_60/gr_MEC018v010101p.pdf, [Online;
accessed 23-Jan-2019].
- [2] 张建敏, 杨峰义, 武洲云等. 多接入边缘计算 (MEC) 及关键技术[M]. 北京: 人民邮电出版社, 2019: 196-217.
- [3] 王秋宁, 谢人超. 移动边缘计算的移动性管理研究[J]. ZTE TECHNOLOGY JOURNAL, 2018.
- [4] 陈建业. 移动边缘计算环境下的服务迁移策略设计与实现[D]. 北京: 北京邮电大学, 2018.
- [5] 黄祥岳. 移动边缘计算缓存优化与用户移动性预测研究. MS thesis. 浙江大学, 2018.
- [6] 刘海鹏. 一种基于差异化服务的 MEC 应用实例智能迁移机制[J]. 电信科学, 2018, 34(9): 118-129.
- [7] Balasubramanian, Venkatraman, et al. "A Mobility Management Architecture for Seamless Delivery of 5G-IoT Services." ICC 2019-2019 IEEE International Conference on Communications (ICC). IEEE, 2019.
- [8] Solozabal R, Sanchoyerto A, Atxutegi E, et al. Exploitation of mobile edge computing in 5G distributed mission-critical push-to-talk service deployment[J]. IEEE Access, 2018, 6: 37665-37675.
- [9] Wang S, Urgaonkar R, Zafer M, et al. Dynamic Service Migration in Mobile Edge Computing Based on Markov Decision Process[J]. IEEE/ACM Transactions on Networking, 2019.
- [10] Kaneda J, Arakawa S, Murata M. Effects of Service Function Relocation on Application-level Delay in Multi-access Edge Computing[C]//2018 IEEE 5G World Forum (5GWF). IEEE, 2018: 399-404.
- [11] Zhang Y, Li J, Zhou Z, et al. Efficient dynamic service maintenance for edge services[J]. IEEE Access, 2018, 6: 8829-8840.

6 网络能力开放

6.1 概述

MEC 网络能力开放 (Network Capability Exposure) 主要包括网络能力的开放功能 (Network Capability Exposure Function) 以及网络服务能力的开放 (Network Service Capability Exposure)。在 MEC 架构体系中, ETSI 标准中定义 MP1 接口的参考功能, 主要包括如服务注册、服务发现和服务通信等功能。同时还提供 MEC 应用可用性检查、会话状态迁移、流量规则和 DNS 规则配置激活更新, 以及存储和时间信息等服务。基于 MP1 接口, MEC 应用可以获取并使用网络能力和网络服务能力^[1]。参考如下图 4-6-1, 网络功能开放基础框架。

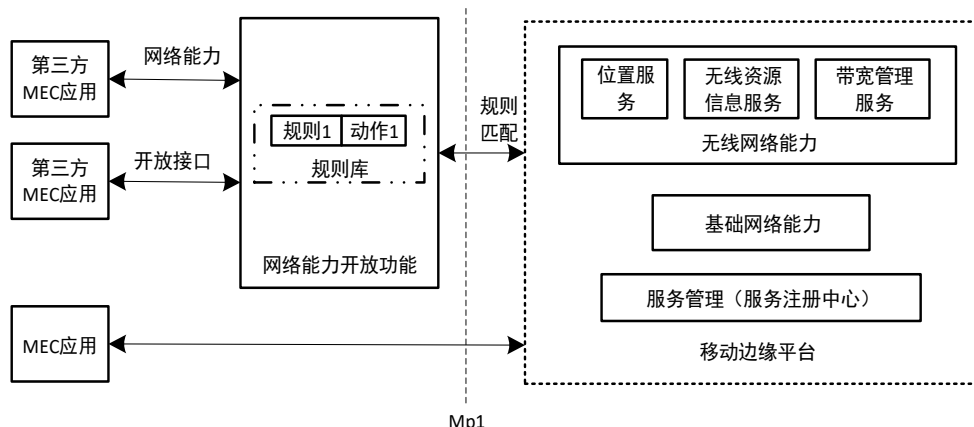


图 4-6-1 网络能力开放基础框架

如上图所示，MEC 应用可以与注册的服务以面向服务的体系结构（Service-Oriented Architecture, SOA）进行通信交互。网络能力开放功能，为 MEC 应用提供标准的抽象接口之外，还能实现对 MEC 应用接口的适配。网络开放功能中包含规则库，规则库完成一序列规则匹配和相应执行动作。无线网络能力针对无线蜂窝网络环境，主要包括无线资源信息服务、位置服务、以及带宽管理服务。MEC 业务需要在 MEC 服务内注册，以及支持面向第三方的发现功能。根据有关注册信息，第三方可以获知某个具体平台业务的状态（例如可用性）以及相关的通信接口。

MEC 网络接入能力的从无线蜂窝网络扩展到 Wi-Fi、固网等环境，基础网络能力在提供多接入网络类型的基础网络能力，如 DNS、路由、交换、DPI 等。网络能力开放的内容主要包括：网络及用户信息、业务及资源控制，以及网络业务处理。随着网络功能的完善和丰富，开放的内容也在不断增加^[2]。通过 MEC 网络能力开放，可以面向第三方提供网络能力开放的内容和能力，实现多方共赢。

（1）网络及用户信息

网络及用户信息包括实时或非实时的网络及用户信息。其中实时信息包括：用户位置、蜂窝负载信息、链路质量统计（CRI、SINA、BLEP）及网络吞吐量统计等。非实时信息包括终端能力和数据连接类型，用户签约信息（如签约标识及优先权）等。

（2）业务及资源控制

业务及资源控制能力指对业务质量、网络运行策略和安全方面的控制和监控功能。例如，通过路由策略控制，第三方应用可根据业务需求，灵活配置数据转发功能，实现路由定制。对于 MEC 边缘计算而言，第三方应用可根据交付内容，通过路由定制，确定将业务流量卸载至边缘云还是私有云，或者按照正常方式通过移动核心网络回传至其集中化的数据中心。

（3）网络业务处理

网络业务处理指包括语音、短消息及多媒体消息服务在内的传统通信业务。

6.2 3GPP SCEF

3GPP TR 23.708 基于移动网络业务能力开放的场景、需求、能力开放架构和关键技术等几方面考虑，对网络能力开放架构进行了设计。为满足应用场景需求，3GPP R13 定义了能力开放架构，引入业务能力开放网元 SCEF（Service Capability

Exposure Function，网络能力开放功能），并定义了 SCEF 相关的网元接口以安全开放网络业务能力，SCEF 可实现的主要功能为将核心网网元能力开放给各类业务应用，通过协议封装及转换实现与合作/自有平台对接，SCEF 的应用使网络具备了多样化的运营服务能力。

如图 4-6-2 所示，SCEF 架构基于网络业务 API，并提供发现功能，帮助第三方快速定制开发个性化的电信业务。通过各类接口可开放的网络能力主要包括：通信信息获取、位置信息获取、通信服务开通信息获取、网络设置信息获取等。API 规范可由 OMA、GSMA 或其他标准组织制定。SCEF 的引入使得能力开放网络配置简单、高效。通过对单个网元 SCEF 进行配置，即可实现核心网多网元各种能力的开放，而不需要在每个核心网网元中新增能力开放模块；同时，引入 SCEF 网元对接各种业务需求的合作/自有平台，可自动将平台需求参数通过 SCEF 自动下发至核心网网元，避免了网络维护人员工作量的增加以及人为配置可能出现差错的风险，可确保网络配置的正确率及网络安全，提高网络运维效率。

基于 SCEF 网络架构定义，可开放的网元能力及应用主要包括下面几点。

- (1) 对 PCRF (Policy and Charging Rules Function，策略与计费规则功能网元) 的能力实现开放
- (2) 对 HSS/MME (Home Subscriber Server/ Mobility Management Entity，归属用户服务器/移动性管理实体) 实现能力开放
- (3) 对更多核心网能力实现开放

面向未来 5G 应用的多种业务场景及多样化用户需求，运营商还可将更多的网元能力进行开放。例如开放 S-CSCF (Serving-Call Session Control Function，服务-呼叫会话控制功能) 与 SCEF 之间的 ISC (IP Multimedia Subsystem Service Control，IMS 业务控制) 接口可实现多方会议、号码显示等语言类能力的开放等，并最终可基于 4G/5G 融合架构通过 NEF (Network Exposure Function，网络功能开放) 与 SCEF 统一的北向接口实现更多核心网网元能力的融合开放。

SCEF 的主要功能如下：

- (1) 鉴权及授权；
- (2) API 调用请求识别；
- (3) 概要管理；
- (4) 接入控制管理；
- (5) 面向外部实体的开放能力发现功能；
- (6) 策略管理包括基础设施策略，例如防止短消息服务中心过载；业务策略，例如业务路由设置；应用策略，例如流量限速。
- (7) 流量记账功能；
- (8) 外部连接功能；
- (9) 网络能力抽象，主要指屏蔽底层 3GPP 网络的接口及协议，支持网络功能集成。具体来说，网络功能集成包括底层协议连接，路由及流量控制，将 API 映射至对应的网络接口、协议翻译。

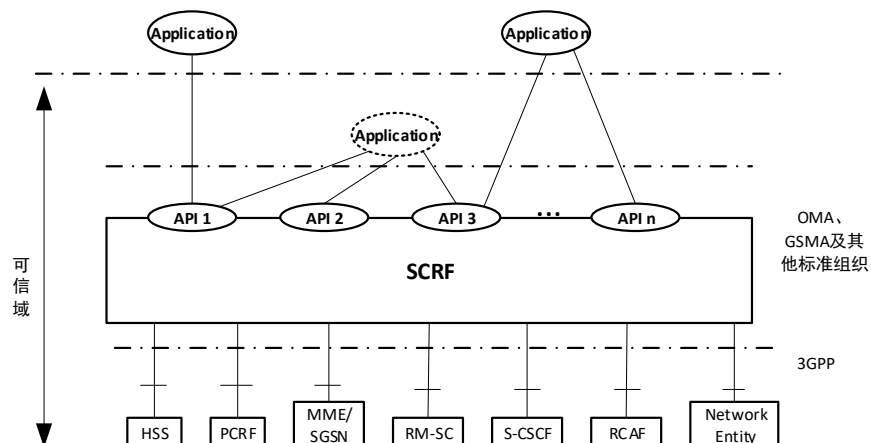


图 4-6-2 网络能力开放基础框架

5G 架构规范 TS 23.501 中将 NEF 作为架构中的重要功能，如图 4-6-3 所示，且控制功能集中部署、切片化的架构设计更利于网络能力开放。另外，5G 架构的网络还可将虚拟化管理及编排能力等新型网络能力对外开放，实现业务自主编排、网络服务及功能差异化定制、网络资源按需伸缩等多类型能力开放应用。NEF（Network Exposure Function，能力开放功能）：主要负责网络能力的收集、分析和重组；完全前提下的网络能力开放。

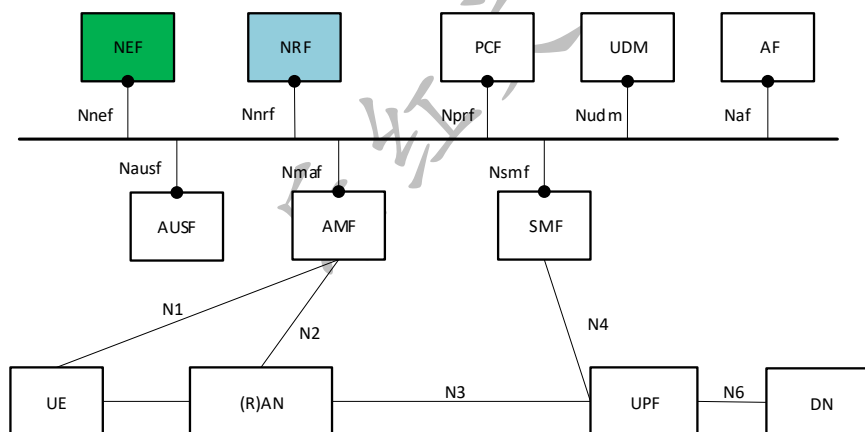


图 4-6-3 基于 5GC 网络架构的能力开放

6.3 API 标准综述

针对无线网络能力开放 ETSI MEC 标准组织制定了相应的 API 接口规范：GS 012，描述了无线网络信息的接口规范；GS 013，描述了位置信息服务的接口规范、GS 014，描述了 UE 身份标识的接口规范、GS 015，描述了带宽管理服务的接口规范。

➤ RNIS

无线网络信息服务（Radio Network Information Service，RNIS）可以提供给授权的 MEC 应用，并可在 Mp1 接口发现服务。通过 RNIS，MEC 应用可以获取无线网络相关信息，主要包括：

- (1) 关于无线网络状况的最新无线网络信息；

- (2) 基于 3GPP 规范的与用户面相关的测量信息；
- (3) 连接到与 MEC 主机相关联的无线节点的 UE 的信息，包括 UE 上下文和相关无线承载信息；
- (4) 与连接到 Internet 的信息有关的变更；

RNIS 可由 MEC 应用程序和平台使用，以优化现有应用，并提供基于无线最新信息的新类型服务。其中，基于无线网络信息优化的典型例子就是移动视频加速，通过获取无线网络可用带宽参考信息并反馈给视频服务器及时调整数据发送速率，既避免了由网络拥塞误判导致的速率下降，尤其是视频观看体验下降问题，同时降低了网络拥塞出现的频次。无线网络信息服务 API，服务使用者可以使用这些 API 来获取无线网络上下文和状态信息，根据这些信息及时反向调整业务内容。

目前，无线网络信息 API 支持查询和订阅（发布/订阅机制）RESTful API 或 MEC 平台的消息代理、能力开放接口采用 HTTP/HTTPS 标准协议，并遵循互联网业界通行的 REST（Representation State Transfer，表现层状态转换）接口设计风格。无线网络信息服务开放的资源 URL（Uniform Resource Locator，统一资源定位符）包含两部分 API root 和自定义资源结构；API root 结构如下：

//{apiRoot}/{apiName}/{apiVersion}

名称	说明
apiRoot	通常为主机名+端口+基本路径，服务的消费方通过服务发现过程从 SR 获得。也可通过订阅接口获得，即服务消费方上线后，订阅该服务
apiName	通常为该服务名，通过服务发现过程获得，对于无线网络信息服务，名称固定为{rni}，即为{apiRoot}/{rni}/{apiVersion}/
apiVersion	使用的 API 版本（当前 ETSI 定义为{V1}）

无线信息查询 API 主要包括：

- (1) 无线 RAB（Radio Access Bear，无线接入承载）信息，主要标识无线侧 RAB 承载的主要参数和状态；
- (2) S1 承载信息，主要标识 S1 侧承载的传输和 QoS 信息；
- (3) PLMN（Public Land Mobile Network，公共地面移动网络）信息，主要标识当前小区的 PLMN 信息；

订阅、注销订阅 RNI 事件 API，订阅和取消订阅 RNI 事件通知是服务创建者需要实时获得无线网络状态和状态变更，并在一定的场景下，需要这些事件变更作为触发源。例如，进行无线网络优化变化时，可以根据特定终端的持续性小区切换或者位置更新通知进行数据采集，获得统计分析值。主要包括如下 API：

- (1) 小区变更事件通知 API；
- (2) RAB 的建立、更新、释放等事件通知 API；
- (3) UE 测量报告通知 API；

(4) 载波聚合重配通知 API;

(5) S1 事件通知等;

➤ LS

位置服务 (Location Service, LS) , 在 ETSI 中定义的 Mp1 接口上注册和发现。目前, 支持位置检索基站, 即每个位置仅报告一次位置信息。同时, 支持位置订阅机制, 即对每个位置请求可多次报告, 或者周期性地, 或者基于特定事件, 例如位置改变, 报告位置信息。除此之外, 位置服务支持匿名位置报告, 即在没有相关的 UEID 信息, 例如统计集合。通常的位置服务既可以支持地理位置 (例如地理坐标), 也可支持逻辑位置 (例如小区)。目前, 支持下列位置信息:

- (1) 与 MEC 主机相关联的无线基站, 当前服务的特定 UE/所有 UE/特定种类 UE 的位置信息
- (2) 某一特定位置的 UE 列表
- (3) 出/入某一特定位置的特定 UE;
- (4) 当前与 MEC 主机相关的所有无线基站的位置信息;

与 rni 类似, API 格式为{apiRoot}/location/{apiVersion}/。主要 API 接口包括:

- (1) 获取某个特定 UE 的位置信息 API;
- (2) 获取某个位置下 UE 所有信息的 API;
- (3) UE 位置查询、订阅、取消 API;
- (4) UE 信息查询、订阅、取消 API;
- (5) 无线基站位置查询 API

位置服务使用者通过 LS API 与 MEC 平台提供的定位服务进行通信, 位置服务 API 支持查询和订阅 (发布/订阅机制), 可在 RESTful 上使用 OMA (Open Mobile Alliance)API 中机制。

➤ UEID

UE 身份 (UE Identity, UEID) , 目的是允许 MEC 系统中的 UE 特定业务规则。MEC 平台提供了支持 UE 身份特征标识功能, MEC 应用程序可以注册标签 (代表 UE) 或者标签列表。每个标签都映射到一个特定的移动网络运营商系统中的 UE, MEC 平台可以提供这些映射信息服务。

所有资源 URL, {apiRoot}/ui/v1/。API 接口包括:

- (1) UE 身份标签的注册和取消注册;
- (2) UE 身份标签的检索或查询;

➤ BWMS

带宽管理服务（Bandwidth Management Service, BWMS），如下图 4-6-4 所示，BWMS 服务作为 MEC 平台能力的一部分，为不同的 MEC 应用程序或 UE 会话（如 CDN）提供不同的带宽要求（如带宽大小、优先级）。

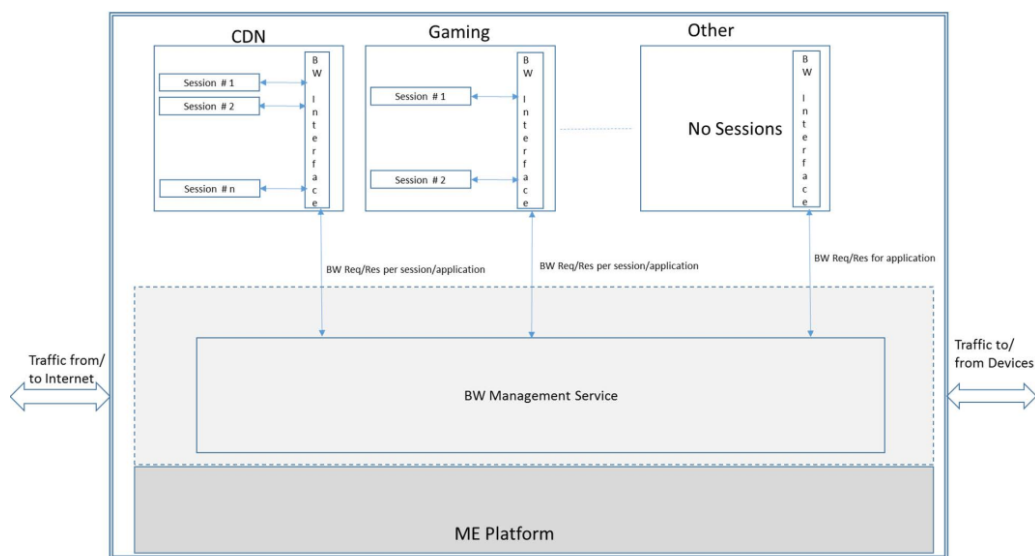


图 4-6-4 带宽管理服务描述

在同一个 MEC 主机上运行的不同 MEC 应用程序可能需要特定的静态/动态上下行带宽资源，包括带宽大小和优先级。在某些情况下，不同的会话在同一应用程序上并行运行可能各自具有特定的带宽要求。除此之外，UE 会话也会导致带宽资源的变更。所有这些应用程序和应用程序会话在相同的共享带宽资源上镜像，带宽资源管理服务可以满足这些特殊要求。

BWMS 主要功能包括：

- (1) 注册和取消注册带宽管理服务；
- (2) 更新所请求的带宽管理服务；
- (3) 获取所分配的带宽分配信息；

主要 API 接口包括：

- (1) 带宽资源的申请分配；
- (2) 带宽资源列表的获取或查询；

6.4 参考文献

- [1] ETSI. Mobile edge computing-introductory technical white paper[R]. 2014..
- [2] 俞一帆，任春明，阮磊峰等著，5G 移动边缘计算[M].北京：人民邮电出版社，2017:71-72.

7 计算卸载

7.1 概述

计算卸载 (Computation Offloading) 通常是需要某个计算量大的任务根据一定的卸载策略合理分配给资源充足的远程设备处理。一般也称为计算迁移 (Computation Migration)，或者任务卸载 (Task Offloading)。由于移动终端的计算资源、存储资源和电池容量有限，可能无法满足某些应用对处理能力、续航能力等的需求。传统意义上的计算卸载是指某个计算过程从移动设备迁移到有更丰富资源的云端上，利用云端收集、存储和处理数据，从而减少移动设备的程序执行时间并降低其能耗，如下图 4-7-1 所示，传统的集中式计算模式中，整个终端上完成应用程序的执行；而分布式计算模型中，移动终端通过计算卸载将应用程序任务迁移至云端执行。例如，计算密集型或交互密集型的应用（如，3D 赛车、浮点性能测试，车联网、人脸识别、视频优化等）。

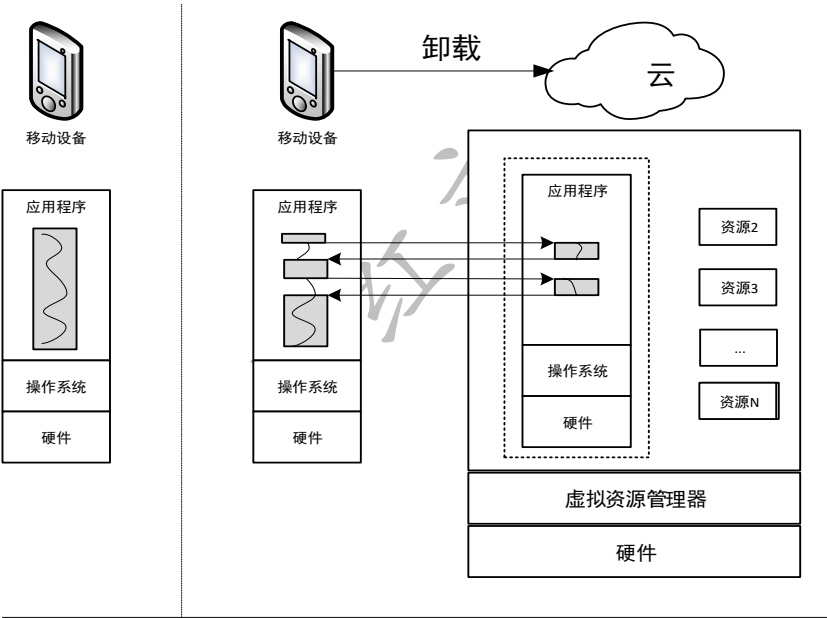


图 4-7-1 计算卸载基础原型

MEC 计算卸载技术是指受资源约束的设备完全或部分地将计算密集型任务卸载到附近资源充足的 MEC 主机上，主要解决了移动设备在资源存储、计算性能以及能效等方面存在的不足。MEC 计算卸载技术不仅减轻了核心网的压力，而且降低了因传输带来的时延。很明显，在传统的云计算中，用户通过访问 Internet 使用云服务。而在 MEC 中，计算和存储资源分布于用户的近距离网络拓扑范围内。与云计算相比，MEC 可以明显降低延迟和抖动，但是 MEC 只能够提供有限的计算和存储资源。云计算和 MEC 的模型差异比较见下表 4-6-1。

表 4-7-1 云计算与 MEC 的模型差异比较

评价指标	云计算	MEC
服务硬件	大型数据中心，有大量高性能计算服务器	小型数据中心，中等有限计算资源

部署方式	集中式	分布式
部署位置	大型数据中心侧，核心网传输后端	5G/LTE 基站处、Wi-Fi 路由侧、其他接入侧网关、核心网传输前端等
与用户的距离	远	近
延迟	高 (>100ms)	低 (<10ms)
抖动	大	小
位置感知	不适用	支持
应用	是时延敏感型应用，如 AR、自动驾驶、交互式在线游戏等	对时延要求较低应用，如社交网络、在线教育、商业购物等

7.2 主要流程

MEC 系统模型计算卸载的主要系统目标是在满足用户延时需求的同时最小化移动设备的总体能耗。基于 MEC 卸载带来优势的同时也提高卸载技术复杂度，需要考虑诸如移动性、资源协同、虚拟环境实时性、无线网络切换等动态或静态的 MEC 环境影响。静态环境不考虑复杂环境变化情况下 MEC 卸载主要的关键流程，参考图 4-7-2，主要包括代理（MEC 服务主机）发现、任务划分、卸载决策、任务提交、任务执行、计算结果反馈^[1]。

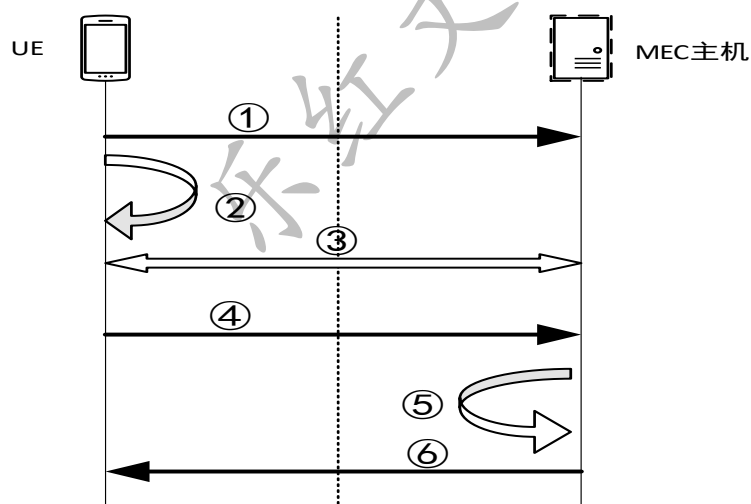


图 4-7-2 MEC 计算卸载主要流程

(1) 代理发现

UE 终端根据自身的数据类型或者任务类型在 MEC 网络环境中找到合适的 MEC 主机服务器，在此服务器上能够完成计算卸载任务注册，MEC 主机服务为该终端提供一个代理服务会话并可能需要创建一个任务实体。

(2) 任务划分

按照预先设置好的算法将一个计算密集型的任务划分为本地执行部分和 MEC 远端执行部分。同时，按照计算任务的细粒度划分出基于进行或功能函数的细粒度（部分）卸载任务或基于应用和虚拟机的粗粒度（全）卸载任务。将这些任务抽象成可执行单元，并按照预设算法同时卸载到多个不同执行功能的代理 MEC 主机上执行。

(3) 卸载决策

计算卸载的过程中会受到不同因素的影响，如用户的使用习惯、无线电信道的通信情况、回程连接的质量、移动设备的性能和云服务器的可用性等，因此计算卸载的关键在于指定适合的卸载决策。在卸载系统中，UE 一般由代码解析器、系统解析器和决策引擎组成，其执行卸载决策分为 3 个步骤：首先，代码解析器确定什么可以卸载，具体卸载内容取决于应用程序类型和代码数据分区。然后，系统解析器负责监控各种参数，如可用代理性能、用户偏好、带宽、网络质量、要卸载的数据大小或执行本地应用程序所消耗的能量；最后决策引擎确定是否卸载，利用计算卸载算法择优完成整个卸载决策过程。

(4) 任务提交

将决策好的可执行单元或任务，通过 4G/5G/Wi-Fi 等网络提交给 MEC 主机处理。

(5) 任务执行

MEC 主机上的创建 MEC 应用实例或使用驻留 MEC 应用执行任务处理。MEC 环境中，UE 终端网络环境或者接入网络类型的变化，必须考虑 UE 移动性给卸载任务带来的卸载失败问题。

(6) 结果反馈

MEC 主机上的 MEC 应用实例将计算结果通过网络回传给 UE 终端。

7.3 关键技术

目前，在基于 MEC 的计算卸载领域中，研究的关键问题主要有 4 个，分别是卸载决策、资源分配、卸载系统和移动性管理^{[2][3]}。

1) 卸载决策

卸载决策是指 UE 决定是否卸载、卸载多少以及卸载什么的问题。卸载决策结果分为本地执行、完全卸载和部分卸载 3 种形式。

- 本地执行 (local execution)：整个计算在 UE 本地完成；
- 完全卸载 (full offloading)：整个计算由 MEC 卸载和处理；
- 部分卸载 (partial offloading)：计算的一部分在本地处理，而另一部分则卸载到 MEC 服务器处理。

目前，计算卸载的性能通常以时间延迟和能量消耗作为衡量指标。具体决策结果由 UE 能量消耗和完成计算任务时延决定，具体分为以下两种情况。

1) 在不进行计算卸载时，时间延迟是指在移动设备终端处执行本地计算所花费的时间；能量消耗是指在移动设备终端处执行本地计算所消耗的能量。

2) 在进行计算卸载时，时间延迟是指卸载数据到 MEC 计算节点的传输时间、在 MEC 计算节点处的执行处理时间、接收来自 MEC 计算节点处理的数据结果的传输时间三者之和；能量消耗是指卸载数据到 MEC 计算节点的传输耗能、接收来自 MEC 计算节点处理的数据结果的传输耗能两部分之和。

卸载方式主要基于可分配到资源大小（包括 MEC 主机资源和无线网络资源）、计算和回传时间的时延，以及完成计算的功耗大小来进行决策。卸载需要考虑的可能因素：

- 网络连接：传输技术（4G、5G、Wi-Fi）、带宽、回传延时、信道状态等；
- 移动设备：CPU 处理速度、内存存储、电池容量和性能等；
- 用户：网络数据传输开销、MEC 主机开销、数据隐私、应用程序执行速度、电池能耗等；
- 应用程序：可卸载性、数据可用性、传输和待处理的数据量、任务划分粒度等；
- MEC 主机服务器：计算处理能力、内存、存储、运行支撑环境、调度能力和资源使用情况等。

目前的卸载决策方案主要包括：

- 以降低时延为目标的卸载决策
- 以降低能量消耗为目标的卸载决策
- 权衡能耗和时延为目标的卸载决策

2) 资源分配

一旦完成了卸载决策，接下来就要考虑合理的资源分配的问题，即卸载在哪里的的问题。如果 UE 的计算任务是不可分割的或可以分割但分割的部分存在联系，这种情况下卸载任务就需要卸载到同一个 MEC 主机。而对于可以分割但分割的部分不存在的计算任务，则可以将其卸载到多个 MEC 主机。考虑的资源包括：

● 无线资源分配

针对复杂的无线通信系统，如何有效分配通信资源，优化无线传输速率，降低用户之间的干扰，提高任务卸载效率。

● 计算资源分配

根据任务的不同计算需求，如何高效分配计算资源，实现 MEC 计算资源的最大化利用。

目前计算资源分配节点主要分为：

● 单节点分配

单节点是指一个基站（MEC 主机与基站绑定）在一个时隙内只能服务于一个计算任务。针对 MEC 网络场景，单节点计算资源分配方案主要是借助于基站间的迁移进行资源分配，如图 4-7-3，用户 1 的计算任务已经分配到基站 1，同样距离基站

1 较近的用户 2 只能将计算任务卸载到附近的基站 2 或基站 3，由于用户 2 卸载给基站 3 传输时延较小，顾选择基站 3 进行计算卸载。

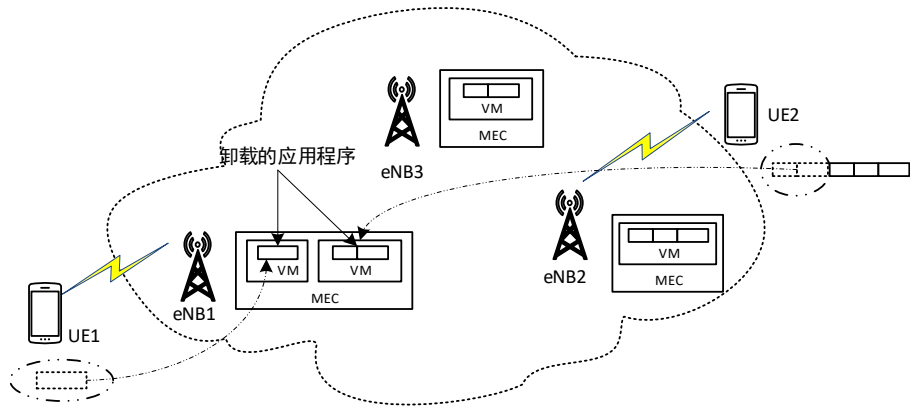


图 4-7-3 基于基站间迁移的单节点计算资源分配

● 多节点分配

多节点是指一个基站（MEC 主机与基站绑定）在一个时隙内可以服务于多个计算任务。如图 4-7-4 所示，用户 1 和用户 2 将计算任务卸载给基站 1 处的 MEC 主机上执行，而此时基站 1 处 MEC 的计算资源已经占满，附近用户 3 将计算任务划分定位两部分卸载给附近的基站 2 和基站 3 处的 MEC 主机，在基站 2 处的 MEC 主机与用户 4 和用户 5 共享计算资源，在基站 3 处的 MEC 主机与用户 6 共享计算资源。

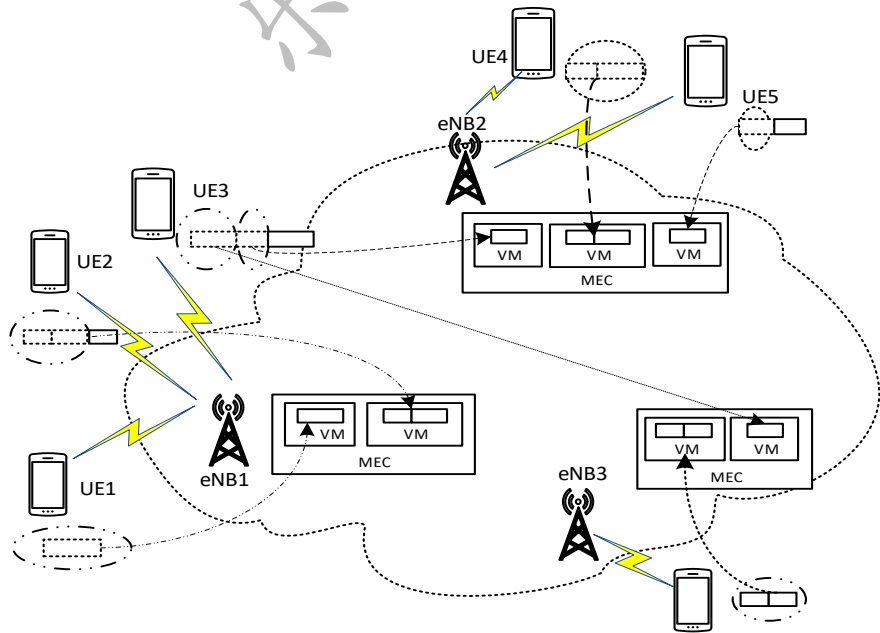


图 4-7-4 基于基站间迁移的多节点计算资源分配

● 协作式节点分配

在多节点之间的计算资源分配中，大部分只是致力于节点之间的合理资源分配，在实现计算卸载时，在各节点之间没有设置一致性目标来优化整个卸载方案，协作式节点分配目标是以一致性目标来优化整个 MEC 计算卸载，以采用一致性算法来解决 MEC 的资源协同问题。

3) 卸载系统

为终端和 MEC 主机提供统一计算卸载系统模型，或者计算卸载编程范式。如 MAUI、Cloudlet、CloneCloud 计算卸载模型，其中 CloneCloud 通过使用静态代码分析和动态环境分析相结合的方法，对应用代码进行划分^[4]。卸载模型主要内容包

4) 移动性管理

针对 UE 用户的移动性，移动蜂窝网络通过集中间的切换来保证服务的连续性和 QoS。类似的，如果 UE 用户将计算任务卸载给 MEC 主机，UE 此时从一个基站切换到另一个基站，而 MEC 主机可能已经不是同一个主机，为保证计算卸载服务的连续性，在考虑 MEC 卸载任务时必须考虑移动性环境对卸载任务的影响问题。

● 基于功率控制的移动性管理

主要针对用户移动性较低的场景，在应用卸载该 MEC 处理过程中，自适应地动态调整基站功率来保证服务的连续性和不中断。功耗调整模式大致分为两种：粗条模式下，基站之间将功率到一个提前设定好的门限值；细调整模式，基站根据实时的情况逐步地将功率调整到一个合适的水平。

● 基于虚拟机或虚拟镜像迁移的移动性管理

如果用户切换到了一个新的服务基站，那么服务的连续性将通过计算节点间的虚拟机迁移得到保障，并将用户和计算节点之间选择一条合适的路径。虚拟化分为，系统级虚拟化和容器虚拟化。虚拟化迁移主要考虑两个指标。一个是迁移开销 $Cost_M$ ，代表着虚拟镜像迁移所带来的核心网资源损耗和时延损耗；另一个迁移增益 $Gain_M$ ，代表着虚拟镜像给用户端带来的时延降低和用户计算结果回传资源的节约。如果迁移开销 $Cost_M$ 大于迁移增益 $Gain_M$ ，则不执行本次迁移；反之，则执行。如下图所示，基站 1 处计算进行需要进行虚拟迁移，如果卸载到基站 2 或 3 的迁移开销 $Cost_M$ 大于迁移增益 $Gain_M$ ，故不选择。待搜索到基站 n 时，得到迁移开销 $Cost_M$ 小于迁移增益 $Gain_M$ ，于是选择基站 n 进行计算卸载。

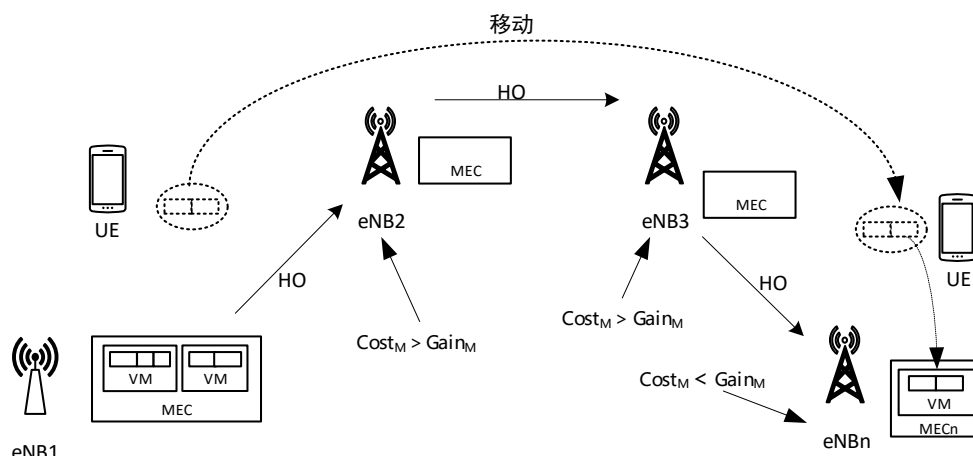


图 4-7-5 虚拟机迁移基本过程

7.4 研究进展

MEC 计算卸载技术成为移动网络研究的热点之一。围绕 MEC 计算卸载的研究，既要考虑设备侧与 MEC 环境侧的动态/静态影响，也要考虑应用程序业务类型差异对卸载决策的影响。计算卸载研究的同时还会协同考虑其他技术影响，如 MEC 服务部署、资源能耗管理、移动性管理、缓存数据加速等。本文罗列出部分 MEC 计算卸载研究进展，如下表所示描述。

刊文	内容概述
移动边缘计算卸载技术综述 ^[3]	首先介绍了 MEC 的网络架构及其部署方案，并对不同的部署方案做了分析和对比；然后从卸载决策、资源分配和系统实现这 3 个方面对 MEC 计算卸载关键技术进行了研究；通过对 5G 环境及其 MEC 部署方案的分析提出了两种计算卸载优化方案，总结归纳了目前 MEC 中计算卸载技术面临的移动性管理、干扰管理以及安全性等方面的核心挑战。
移动边缘计算中的计算卸载和资源管理方案 ^[6]	总结现有 MEC 资源管理的最新研究成果，从基于二进制计算卸载的计算卸载方案、基于部分卸载计算的计算卸载方案、基于随机任务模型的计算卸载方案和云-MEC 协同的计算卸载方案 4 个方面介绍了计算卸载决定、无线资源、计算资源联合优化的计算卸载算法。
Mobile edge computing: A survey on architecture and computation offloading ^[7]	文中首先描述了适用 MEC 的主要用例和参考方案。然后，调查并讨论了 MEC 标准化的最新进展。该调查的核心集中在 MEC 计算分流用例上。同时，将计算卸载的研究分为三个关键领域：1) 计算卸载决策；2) 在 MEC 计算资源分配；3) 移动性管理。最后，重点介绍了 MEC 领域的研究难点和未来发展方向。
Joint Computation Offloading and Resource Allocation Optimization in Heterogeneous Networks With Mobile Edge Computing ^[8]	提出了一种异构网络中的移动边缘计算的分布式联合计算卸载和资源分配优化 (JCORA) 方案。提出了一个优化问题，给出了最优计算卸载策略、上行子信道分配、上行传输功率分配和计算资源调度。为了联合分配上行链路子信道、上行链路传输功率和卸载移动终端的计算资源，设计了云与无线资源分配算法。分布式 JCORA 方案，通过两个子算法的相互迭代来解决优化问题。仿真结果表明，该方案可以有效地降低能耗和任务完成时间，且具有较低的复杂度。
A cooperative partial computation offloading	物联网协同计算卸载算法，从单用户到多用户之间的资源竞争，将多用户计算卸载问题归结为一个混合整数线性规划问题，这是 NP-hard 问题。由于所提出的问题的计算复杂

scheme for mobile edge computing enabled Internet of Things ^[9]	度, 设计了一种迭代启发式 MEC 资源分配算法来动态地进行卸载决策。仿真结果表明, 该算法在执行延迟和卸载效率方面优于现有的方案。
刊文	内容概述
Energy Efficiency Based Joint Computation Offloading and Resource Allocation in Multi-Access MEC Systems ^[10]	一种多址 MEC 服务器系统, 其中正交频分复用接入 (OFDMA) 被用作上行链路的传输机制。为了最小化 MDS (mobile devices) 的能耗, 提出了一种计算卸载、子载波分配和计算资源分配的联合优化策略, 即混合整数非线性规划 (MINLP) 问题。首先, 设计了一个边界改进的分支定界算法 (BNB) 来寻找全局最优解。然后, 提出了一种组合算法, 以获得实际应用的次优解。仿真结果表明, 该组合算法在节能方面与 BNB 算法非常接近, 但在平均算法时间上具有更好的性能。此外, 提出的解决方案优于其他基准方案。
刊文	内容概述
Delay Minimization for NOMA-MEC Offloading ^[11]	考虑了非正交多址辅助移动边缘计算 (nonorthogonal multiple access assisted mobile edge computing, NOMA-MEC) 的卸载延迟的最小化。通过将延迟最小化问题转化为分式规划的形式, 提出了两种基于 Dinkelbach 和 Newton 方法的迭代算法。证明了这两种方法的最优性, 并对它们的收敛性进行了比较。在此基础上, 建立了 MEC 卸载的三种可能模式: 正交多址 (orthogonal multiple access)、纯 NOMA (pure NOMA) 和混合 NOMA (hybrid NOMA) 的选择准则。
刊文	内容概述
Joint Offloading and Computation Energy Efficiency Maximization in a Mobile Edge Computing System ^[12]	本文提出了一种新的算法来评估移动边缘计算系统的性能。具体来说, 一个称为计算效率的新指标定义为计算数据位数除以相应的能耗。为了及时计算所需的数据, 将本地计算和数据分流两种方案结合在一起, 形成联合计算算法。提出了一个优化问题, 其目的是使具有加权因子的用户之间的计算效率总和最大化。使用迭代和梯度下降方法, 可以有效地解决该问题。仿真结果表明, 该方案优于传统方案。
刊文	内容概述
Computation Rate Maximization in UAV-Enabled Wireless-Powered Mobile-Edge Computing Systems ^[13]	本文研究了一种无人机支持的 MEC 无线供电系统。研究了一种无人机支持的 MEC 无线供电系统的计算速率最大化问题, 在能量收集和无人机速度约束条件下, 分别计算了部分和二进制计算卸载模式。分别提出了两阶段算法和三阶段交替算法来解决所提出的问题。针对二进制计算卸载模式, 提出了用户选择本地计算或卸载计算任务的最优选择方案。仿真结果表明, 提出的资源分配方案优于其他基准方案。结果还表明, 所提出的方案收敛速度快。
刊文	内容概述
Offloading schemes in mobile edge computing for ultra-reliable low latency communications ^[14]	本文研究了 MEC 任务卸载的延迟和可靠性之间的权衡问题。提供了一种框架, 其中用户设备将任务分割成子任务, 并依次将其卸载到多个附近边缘节点 (ENs)。在这个框架中, 制定了一个优化问题, 共同最小化延迟和卸载失败概率。由于所提出的问题是凸的, 分别设计了基于启发式搜索、重构线性化技术和半正定松弛的三种算法, 并通过优化 EN 候选者选择、卸载排序和任务分配来解决问题。数值仿真结果表明, 所提出的算法在 URLLC 中的时延和可靠性之间有很好的平衡。其中, 启发式算法以最小的复杂度来实现延迟和可靠性方面的最佳性能。
刊文	内容概述
Joint Task Offloading and Resource Allocation for Multi-Server Mobile-Edge Computing Networks ^[15]	在本文中, 考虑了 MEC 支持的多小区无线网络, 其中每个基站 (BS) 配备有 MEC 服务器, 其协助移动用户通过任务卸载执行计算密集型任务。研究了协同任务卸载和资源分配的问题, 以最大化用户任务卸载增益, 通过任务完成时间和能量消耗的加权和来度量。所考虑的问题制定为混合整数非线性规划 (MINLP), 涉及联合优化任务卸载决策、上行链路传输功率的移动用户, 以及在 MEC 服务器上的计算资源分配。使用凸和拟凸优化技术解决 RA (资源分配) 问题, 并提出了一种新的启发式算法, 实现次优的解决方案在多项式时间内。仿真结果表明, 算法与最优解密切相关, 它显著地提高了用户对传统方法的卸载效用。
刊文	内容概述

Service-dependent task offloading for multiuser mobile edge computing system ^[16]	不同的场景需要不同的服务，并且任务通常需要在特定的运行时环境中进行计算。然而，很少有研究考虑到这一点。为此，作者提出了一种新的基于服务的多用户 MEC 系统资源受限任务卸载模型。通过确定应该在网络边缘部署哪些服务以及卸载多少任务，构建了收益最大化问题。因此，他们提出了一种迭代优化算法，具有高性能的服务部署和任务卸载问题。最后，数值模拟表明，所提出的算法优于其他方案。
刊文	内容概述
Multiuser Computation Offloading and Downloading for Edge Computing With Virtualization ^[17]	本文研究了在 I/O 干扰的情况下，多用户 MEC 系统中的无线资源和计算资源分配（RCRA）协同问题。具体而言，卸载调度算法两个系统性能指标：总和卸载速率最大化和总和移动能耗最小化。它们被设计为非凸混合整数规划问题，考虑了由计算卸载、结果下载和并行计算所引起的延迟。基于分解方法设计了一套低复杂度的算法，并利用组合优化的经典技术。得到了多用户协同卸载调度算法，控制它们的卸载大小，并划分通信时间和计算时间。综合仿真结果表明，考虑 I/O 干扰可以使卸载可控，效果最优。

7.5 参考文献

- [1] 张建敏, 杨峰义, 武洲云等. 多接入边缘计算 (MEC) 及关键技术[M]. 北京: 人民邮电出版社, 2019: 196-217.
- [2] MACH P, BECVAR Z. Mobile edge computing: a survey on architecture and computation offloading[J]. IEEE Communications Surveys & Tutorials, 2017, PP(99):1-1.
- [3] 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, Vol.39 No.11.
- [4] Chun B G, Ihm S, Maniatis P, et al. CloneCloud: elastic execution between mobile device and cloud[C]//Conference on Computer Systems. 2011.
- [5] Ahmed A, Ahmed E.A survey on mobile edge computing[C]//International Conference on Intelligent Systems and Control, 2016.
- [6] 李邱苹, 赵军辉, 贡毅. 移动边缘计算中的计算卸载和资源管理方案[J]. 电信科学, 2019, 35(3): 36-46.
- [7] Mach P, Becvar Z. Mobile edge computing: A survey on architecture and computation offloading[J]. IEEE Communications Surveys & Tutorials, 2017, 19(3): 1628-1656.
- [8] Zhang J, Xia W, Yan F, et al. Joint computation offloading and resource allocation optimization in heterogeneous networks with mobile edge computing[J]. IEEE Access, 2018, 6: 19324-19337.
- [9] Ning Z, Dong P, Kong X, et al. A cooperative partial computation offloading scheme for mobile edge computing enabled Internet of Things[J]. IEEE Internet of Things Journal, 2018.
- [10] Yang X, Yu X, Huang H, et al. Energy Efficiency Based Joint Computation Offloading and Resource Allocation in Multi-Access MEC Systems[J]. IEEE Access, 2019, 7: 117054-117062.

-
- [11] Ding Z, Ng D W K, Schober R, et al. Delay minimization for NOMA-MEC offloading[J]. IEEE Signal Processing Letters, 2018, 25(12): 1875-1879.
- [12] Sun H, Zhou F, Hu R Q. Joint offloading and computation energy efficiency maximization in a mobile edge computing system[J]. IEEE Transactions on Vehicular Technology, 2019, 68(3): 3052-3056.
- [13] Zhou F, Wu Y, Hu R Q, et al. Computation rate maximization in UAV-enabled wireless-powered mobile-edge computing systems[J]. IEEE Journal on Selected Areas in Communications, 2018, 36(9): 1927-1941.
- [14] Liu J, Zhang Q. Offloading schemes in mobile edge computing for ultra-reliable low latency communications[J]. IEEE Access, 2018, 6: 12825-12837.
- [15] Tran T X, Pompili D. Joint task offloading and resource allocation for multi-server mobile-edge computing networks[J]. IEEE Transactions on Vehicular Technology, 2018, 68(1): 856-868.
- [16] Ni W, Tian H, Lyu X, et al. Service-dependent task offloading for multiuser mobile edge computing system[J]. Electronics Letters, 2019, 55(15): 839-841.
- [17] Liang Z, Liu Y, Lok T M, et al. Multiuser computation offloading and downloading for edge computing with virtualization[J]. IEEE Transactions on Wireless Communications, 2019.

8 多接入管理

8.1 概述

多接入场景中，用户终端可以有多个网络可以选择，例如 WLAN、5G、4G、固网等。如果这些不同接入网络之间是独立的，那么要实现多网协同，就需要一个独立于接入网络的高层网络功能在用户终端和远端服务器之间进行交付。IETF 提出了一种称为“多接入管理服务（Multiple Access Management Service, MAMS）”的 Internet 草案，用户终端和 MEC 服务器之间是有 MAMS 的控制面协议协商业务数据流量的传输路径，用户面是有 MPTCP、GRE 等多链路聚合协议，实现业务在多接入环境下的灵活路由^[1]。

8.2 MAMS

MAMS 是一个可编程的框架，它提供了一种多接入网络中的基于业务需求灵活选择网络路径的机制。MAMS 利用网络智能感知能力获取网络和链路状态信息，并基于这些信息动态的调整业务流量在多个网络路径上的分布以及用户面数据的处理方法。网络路径选择和配置消息通过用户面数据进行传输，因此不会影响底层接入网络的控制面信令协议。

参考文献^[2]给出了 MAMS 框架的需求、解决方案原则、功能框架和协议流程。MAMS 不依赖于任何接入网络类型或用户面协议（TCP、UDP、MPTCP、GRE 等），而是基于客户端和网络能力提高了一种协商和配置用户面协议的方法，是在多接

入场景中对用户面协议的一种补充。MAMS 一个重要目标是确保它与底层接入网络没有任何依赖关系或者依赖关系最小，实际实现上 MAMS 功能元素形成了一个多网络传输路径上的 IP 叠加层。MAMS 提供了轻量级的多接入聚合方案，并支持灵活地功能部署。

图 4-8-1，MAMS 参考架构。控制面功能元素有网络连接管路（Network Connection Manager，NCM）和用户连接管理（Client Connection Manager，CCM）。用户面功能元素包括网络多接入数据代理（Network Multiple Access Data Proxy，N-MADP）和客户端多接入数据代理（Client Multiple Access Data Proxy，C-MADP）。

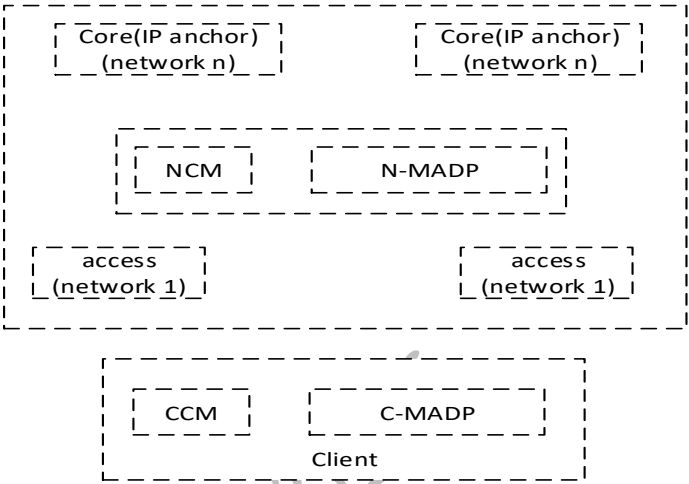


图 4-8-1 MAMS 参考架构

MAMS 控制面协议^[3]提出了一种基于 TCP 的 MAMS 控制面协议栈，参考图 4-8-2。WebSocket 用于在 NCM 和 CCM 之间传输管理和控制消息。TCP/TLS 上承载控制面消息与底层的传输网络无关。多接入控制消息根据应用需求、用户和网络的能力配置用户面协议，并支持根据网络条件动态自适应最优网络路径和用户面协议。

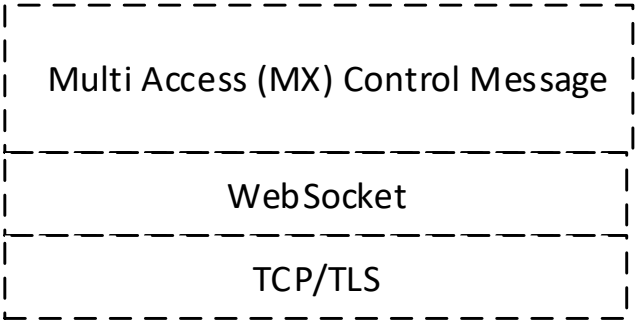


图 4-8-2 MAMS 控制面协议栈

MAMS 控制面协议^[4]给出了 MAMS 用户面控制协议，参考图 4-8-3。用户面协议栈包括：多接入汇聚子层（Multi-Access Convergence Sublayer）和多接入适配子层（Multi-Access Adaptation Sublayer）。

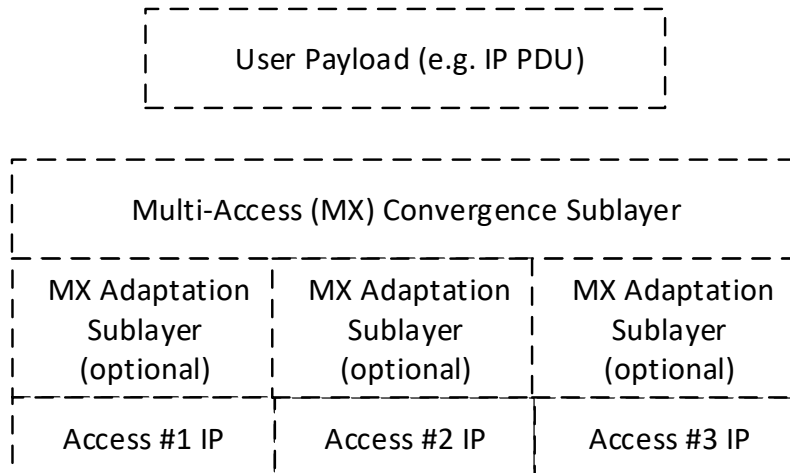


图 4-8-3 MAMS 用户面协议栈

8.3 MAMS 与 MEC

MEC 在接入网络边缘提供了计算平台，MAMS 在 MEC 平台上适用性已在不同的网络配置条件下进行了评估测试。N-MADP 可以作为 MEC 应用与 NCM 一起部署在 MEC 服务器上，也可以部署在别的网元上，例如多个接入网络共用的网关或路由器上。MEC 平台提供了一种可能的 MAMS 部署方案，并且能够通过减少业务时延、开放实时无线网络信息来进一步优化终端用户的业务体验^[1]。

MEC 平台包括网络功能、平台服务和业务应用，参考如下图 4-8-4 所示。MEC 的网络功能负责实现边缘业务应用的数据分组解析与网关出口。具体地，5G 网络中负责 MEC 网络功能的实体是 UPF、WLAN 中负责 MEC 网络功能实体是 BRAS (Broadband Remote Access Server)。多网协同管理的应用需求主要包括以下几种：

- (1) 流量分流：5G 网络的流量分流至城域网，分流策略包括为了减轻移动回传压力的全业务分流、特定业务分流等；
- (2) 流量疏导：根据网络状态、网络策略或本地策略，为业务选择合适的网络；网络状态指的是从用户终端到网络网关之间的端到端链路状态（包括有线和无线链路状态）。
- (3) 流量切换：将业务流量从一个网络切换到另一个网络进行传输，通常是由于第一个网络不可用而导致的；流量切换可以实现网络间的移动性管理。
- (4) 链路聚合：业务可以同时使用多个接入链路进行数据传输，充分利用多接入网络的带宽资源，提升业务体验。

对于流量分流，5G 系统通过 UPF 下沉能够满足该应用需求。而对于流量疏导、流量切换以及链路聚合，则需要专门的多网协同管理功能实现。基于 MEC 的多网协同管理服务 (Multi-Network Management Service, MNMS) 参考架构，如下图 4-8-4 所示。MNMS 技术需求是 MAMS 的子集，因此 MAMS 协议栈完全能够实现 MNMS 所需要的功能。

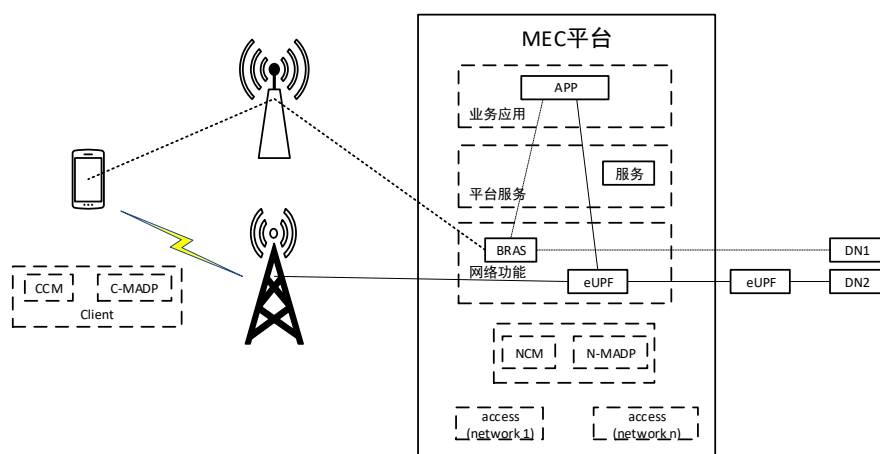


图 4-8-4 MNMS 参考架构

8.4 参考文献

- [1] 张建敏, 杨峰义, 武洲云等. 多接入边缘计算 (MEC) 及关键技术[M]. 北京: 人民邮电出版社, 2019: 196-217.
- [2] S. Kanugovi. Multiple Access Management Service[R].2017.
- [3] J. Zhu. Control-Plane Protocols for Multiple Access Management Service[R].2019.
- [4] J. Zhu .User-Plane Protocols for Multiple Access Management Service[R].2019.

9 资源管理

9.1 概述

随着 MEC 的发展和规模性的部署, MEC 的能耗成为 MEC 服务提供商的重要成本。同时, 数据规模的增长对资源的协同优化提出了更高的要求。如何从 MEC 服务整体上降低能耗, 提高计算、存储和网络的资源利用率, 进而改善网络服务质量和用户体验, 是 MEC 需要解决的问题。MEC 资源管理 (Resource Management, RM) 技术是 MEC 中的关键技术之一, 主要是解决 MEC 系统中计算、存储和网络资源的管理和优化问题。

9.2 关键技术

MEC 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。针对 MEC 环境中计算、存储、通信 3 种关键资源, 本章将从多个技术角度分析如何 MEC 的资源管理和优化。

MEC 资源管理是个从系统性的工程问题, 本章主要从以下几个技术角度分析 MEC 资源管理和优化:

- (1) 面向业务场景的 MEC 主机部署。目前主流的部署是在基站侧和有线网关设备侧采用分布式的方式部署 MEC 服务, 而且每个 MEC 节点都具有一定的计算、存储、网络转发能力, 为连接到该节点的终端提供服务。部署的 MEC 节点需要

考虑的如 MEC 服务节点间的资源协同和资源共享问题、最大化利用 MEC 主机计算、存储、网络资源的，以及如何最大化降低运营商和服务提供者的成本。

(2) 面向用户需求的 QoE 优化。QoE (Quality of Experience, 服务质量体验) 就要求在当前网络环境下提供数据传输速率和提供更低网络延时能力。典型的业务诉求比如视频，基于 MEC 的自适应视频流业务利用自适应比特流 (Adaptive Bit Rate, ABR)、视频转码与视频自适应分发技术 (CDN) 提供更低延时和更高带宽的视频服务。

(3) 面向能效优化的资源管理。MEC 服务框架主要分为用户终端、无线基站/有线网关、核心网、MEC 主机、源服务器这 5 个部分，MEC 服务的能耗均有可能在这 4 个服务或主机侧产生，主要包括数据传输能耗、任务计算能耗、缓存能耗。数据传输能耗主要包括从 MEC 主机到用户之间的无线或有线传输能耗、从源服务器到 MEC 主机之间的回传能耗，以及从 MEC 主机到用户终端之间的回传能耗。任务计算能耗，主要指 MEC 主机执行计算任务产生的能耗，这里可能会产生计算卸载、应用 (或虚拟机) 迁移、以及协同计算等过程。缓存能耗，主要是由于 MEC 服务器缓存内容而造成的能耗，如将数据存储仅缓存空间过程中的能耗、维持缓存所消耗的能量和从缓存中读取数据所消耗的能力等。

(4) 面向资源共享的协作机制。MEC 主机服务一般有多个运营商和服务提供商共同运营，同时 MEC 主机部署一般会采用分布式分级模式部署。相邻的 MEC 服务器之间可以通过协作机制进行缓存和计算，即可以解决缓存内容不足、计算资源超载、单一 MEC 主机所带来的 MEC 服务性能低下、MEC 主机上 CDN 内容缓存内容单一、空闲 MEC 主机资源浪费等问题。MEC 主机节点间的相互协作和资源互惠互利十分有利于 MEC 主机服务质量的提升和资源管理和优化。

(5) 面向资源管理的高性能通信机制。MEC 的多接入网络设备侧多样性和网络接入类型的多样性，针对不同的业务应用场景和网络接入类型，需要解决网络传输和网络传输资源优化这个基本问题。上文中提到的 MAMS 多网协同管理，有效解决多种网络接入场景下业务多样性通信和一致性业务体验，提高带宽资源的利用率。基于 5G 网络中的毫米波的通信技术极大地丰富了网络带宽。以及基于 MEC 设备之间的 D2D 通信技术极大地提高了信道资源的利用效率，提升了 MEC 网络的容量。

(6) 面向现代网络技术的资源有效分配。现代网络技术典型的包括 NFV、SDN、5G 网络切片等。针对不同的业务场景、利用现代网络技术实现资源的高效管理和合理分配也是资源管理和优化的重要内容。NFV 可以为用户提供按需分配的可配置资源共享池，可以极大的方便资源的统一管理。同时 SDN 实现了集中式的控制平面，并可以为用户提供网络编程接口，使得用户可以根据上层业务和应用个性化地定制网络资源来满足其特有的需求。网络切片可以针对不同的应用场景，将物理网络切割成多个虚拟网卡，从而应对不同场景中的传输时延、移动性、可靠性、安全性以及计费方式的差异性，实现业务的多样化的处理。针对不同的业务场景，进行有效的网络业务切片和划分，并通过 SDN 全局管理控制和对 NFV 虚拟化资源合理调度，对优化整体资源效率、网络性能和资源能耗有极好处。

9.3 研究进展

MEC 资源管理和优化是个综合性研究话题，目前主要的研究是面向协作机制的计算卸载、缓存优化、资源全局分配调度，传输优化以及面向用户的 QoE 质量提升。另外还包括针对业务应用场景的资源优化，如车联网、IoT、视频业务等。具体参考，如下表所示描述。

刊文	内容概述
面向视频流的 MEC 缓存转码联合优化研究 ^[1]	本文从 MEC 和 ABR 的背景和概述出发，对目前面向视频流的缓存转码资源联合优化方案进行了介绍和分析，并主要从缓存策略、转码策略和优化方式等方面对已有方案进行了对比。在对以上方案分析对比的基础上，研究了面向视频流的 MEC 资源优化问题目前面临的挑战和研究难点。
刊文	内容概述
多用户移动边缘计算迁移的能量管理研究 ^[2]	针对多用户 MEC 系统中移动设备成本的最小化问题，提出了基于强化学习的资源管理算法，制定了成本最优化情况下移动设备的任务分配策略，并将可再生能源纳入多用户移动边缘系统中，实现了移动设备成本最优化（包括时延成本和能耗成本）。
刊文	内容概述
基于综合信任的边缘计算资源协同研究 ^[3]	本文针对边缘计算的隐私信任和安全以及资源协同优化的关键问题，提出了基于综合信任的边缘计算多层自适应的边缘计算模型与体系结构，并基于综合信任对边缘计算的资源协同优化问题进行了建模，形成了统一的多目标优化解决方案。最后通过边缘计算场景中的流计算任务面向计算、存储以及动态移动的通信资源进行任务调度，优化了资源共享效率，仿真实验验证了方法的有效性，表明该方法能够提升用户的 QoE。
刊文	内容概述
面向多业务需求的 NFV 和 SDN 融合的资源优化算法 ^[4]	本文给出了虚拟网络功能放置与服务功能链路由的联合优化模型，该优化模型是 NP-Hard 问题。为了解决该问题，提出两种启发式算法：一种是先路由后放置（FRTP, first routing then placing）的服务链部署算法，另一种是基于节点优先级的先放置后路由（PFBR, placing followed by routing）的服务链部署算法。仿真结果表明，与其他算法相比，提出的 FRTP 和 PFBR 算法能显著平衡网络中的流量负载，改善链路拥塞情况，从而提高服务功能链请求接受率。
刊文	内容概述
SDN-based resource management for autonomous vehicular networks: A multi-access	本文提出了一种新的网络架构来增强 AVs 之间的协作驱动，在 AVNETs 应用 MEC 将计算和存储资源移动到核心网络的边缘，AVs 通过不同的无线接入技术接入网络。为了提高计算/存储和带宽资源的利用率，提出了一种新的网络架构下的资源管理框架。通过集中管理云计算和 MEC 服务器上的计算/存储资源，可以增强 MEC 服务器之间的任务负载平

edge computing approach ^[5]	衡。此外，每个 MEC 控制器上的 SDN 和 NFV 控制模块还负责在异构 BSs 之间分割不同访问技术的带宽资源，以提高带宽利用率。
刊文	内容概述
LayBack: SDN management of multi-access edge computing (MEC) for network access services and radio resource sharing ^[6]	本文引入了分层回传 (LayBack) 架构，用于将异构无线电接入网络 (RAN) 与软件定义网络 (SDN) 进行协调。LayBack 将异构 RAN 在其各自的网关 (例如小型小区网关的云 RAN (CRAN) 基带单元) 融合。更具体地说，这些异构网关通过 SDN 网络连接到统一的 SDN 编排器。还引入了一个基于 SDN 的管理框架，该框架在 SDN 编排器中执行。管理框架协调分布式计算资源，例如分布式多路访问边缘计算 (MEC) 节点，以内聚地为网络服务提供计算。
刊文	内容概述
Resource management for mobile edge computing using user mobility prediction ^[7]	本文提出了一种使用用户移动性预测的 MEC 资源管理。其中分布式边缘服务器 (ES) 实时执行用户任务。在 MEC 中，ES 是分布式安装的，用户必须决定他们使用哪个 ES 来执行任务，以及考虑 ES 的连通性。当用户任务由多个 ES 处理时，并且当用户从 ES 收集计算结果时，ES 的连接因用户移动性而终止，用户无法收集任务的结果，因此用户必须再次执行任务。由于用户的移动性导致的连接丢失是一个大问题。本文使用卡尔曼滤波器预测用户的移动性，以及连通性预测。使用移动性预测，用户可以在任务请求和任务收集期间选择稳定的 ES。
刊文	内容概述
Energy-Efficient Joint Offloading and Wireless Resource Allocation Strategy in Multi-MEC Server Systems ^[8]	本文中，考虑了一种基于正交频分复用接入 (OFDMA) 的多用户和多 MEC 服务器系统，研究了任务分流策略和无线资源分配。为了最大程度地减少总能耗，提出了针对延迟关键型应用程序的联合卸载和资源分配策略。通过双层优化方法，原始的 NP-hard 问题被分解为寻求功率和子载波分配的下层问题和上层任务卸载问题。仿真结果表明，与传统算法相比，该算法在节能和成功卸载概率 (SOP) 方面具有优异的性能
刊文	内容概述
Joint Radio and Computation Resource Management for Low Latency Mobile Edge Computing ^[9]	提出了一种联合无线电和计算资源管理 (iRAR) 算法，该算法通过优化上行链路传输功率，接收波束成形，计算任务分配和计算资源分配来最大程度地减少用户的服务等待时间。
刊文	内容概述
Efficient Computation Resource Management in	研究了移动边缘云计算网络中的计算资源管理问题。移动边缘服务器应首先满足移动用户和物联网 (IoT) 设备的计算要求，然后将多余的计算资源批发到云网络以最大程度地提高其利润。由于批发的时间粒度较粗，可能偶尔会发生计算资源回购以应对流量突发的

Mobile Edge-Cloud Computing ^[10]	情况。因此，移动边缘服务器需要在批发利润和回购成本之间进行权衡。本文将计算资源管理问题建模为利润最大化。仿真结果表明，我们提出的有效的批发和回购方案可以在保证所有计算任务的计算延迟的同时增加总利润，尤其是在计算工作量随时间变化的情况下。
刊文	内容概述
Energy-Saving Offloading by Jointly Allocating Radio and Computational Resources for Mobile Edge Computing ^[11]	本文中考虑了一个多移动用户 MEC 系统，其中多个智能移动设备（SMD）要求将计算任务卸载到 MEC 服务器。为了最大程度地降低 SMD 的能耗，共同优化了卸载选择、无线电资源分配和计算资源分配。并将能耗最小化问题表示为混合整数非线性规划（MINLP）问题，该问题受特定的应用程序延迟约束。为了解决该问题，提出了一种基于重构线性化技术的分支定界（RLTBB）方法，该方法可以通过设置求解精度来获得最佳结果或次优结果。考虑到不能保证 RLTBB 的复杂性，进一步设计了一种基于基尼系数的贪婪启发式算法（GCGH），通过将 MINLP 问题降级为凸问题来解决多项式复杂性中的 MINLP 问题。许多仿真结果证明了 RLTBB 和 GCGH 在节能方面的增强。
刊文	内容概述
Collaborative mobile edge computing in 5G networks: New paradigms, scenarios, and challenges ^[12]	本文设想了一种实时的，上下文相关的协作框架，该框架位于 RAN 边缘，由 MEC 服务器和移动设备组成，并在边缘融合了异构资源。具体来说，介绍和研究了三个代表性用例，范围从移动边缘编排，协作缓存和处理以及多层干扰消除。展示了拟议方法在促进向 5G 网络演进方面的潜在好处。最后，我们讨论了关键技术挑战和开放研究问题，这些问题需要解决才能有效地将 MEC 集成到 5G 生态系统中

9.4 参考文献

- [1] 李佳, 谢人超, 贾庆民, 等. 面向视频流的 MEC 缓存转码联合优化研究[J]. 电信科学, 2018, 34(8): 76-86.
- [2] 王璐瑶, 张文倩, 张光林. 多用户移动边缘计算迁移的能量管理研究[J]. 物联网学报, 3(1): 73-81.
- [3] 邓晓衡, 关培源, 万志文, 等. 基于综合信任的边缘计算资源协同研究[J]. 计算机研究与发展, 2018, 55(3): 449-477.
- [4] 朱晓荣, 张倩. 面向多业务需求的 NFV 和 SDN 融合的资源优化算法[J]. 通信学报, 2018, 39(11): 54-62.
- [5] Peng H, Ye Q, Shen X S. SDN-based resource management for autonomous vehicular networks: A multi-access edge computing approach[J]. IEEE Wireless Communications, 2019, 26(4): 156-162.
- [6] Shantharama P, Thyagaturu A S, Karakoc N, et al. LayBack: SDN management of multi-access edge computing (MEC) for network access services and radio resource sharing[J]. IEEE Access, 2018, 6: 57545-57561.

-
- [7] Ojima T, Fujii T. Resource management for mobile edge computing using user mobility prediction[C]//2018 International Conference on Information Networking (ICOIN). IEEE, 2018: 718-720.
- [8] Cheng K, Teng Y, Sun W, et al. Energy-efficient joint offloading and wireless resource allocation strategy in multi-MEC server systems[C]//2018 IEEE International Conference on Communications (ICC). IEEE, 2018: 1-6.
- [9] Liu Q, Han T, Ansari N. Joint Radio and Computation Resource Management for Low Latency Mobile Edge Computing[C]//2018 IEEE Global Communications Conference (GLOBECOM). IEEE, 2018: 1-7.
- [10] Zhang Y, Lan X, Li Y, et al. Efficient Computation Resource Management in Mobile Edge-Cloud Computing[J]. IEEE Internet of Things Journal, 2018, 6(2): 3455-3466.
- [11] Zhao P, Tian H, Qin C, et al. Energy-saving offloading by jointly allocating radio and computational resources for mobile edge computing[J]. IEEE Access, 2017, 5: 11255-11268.
- [12] 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.

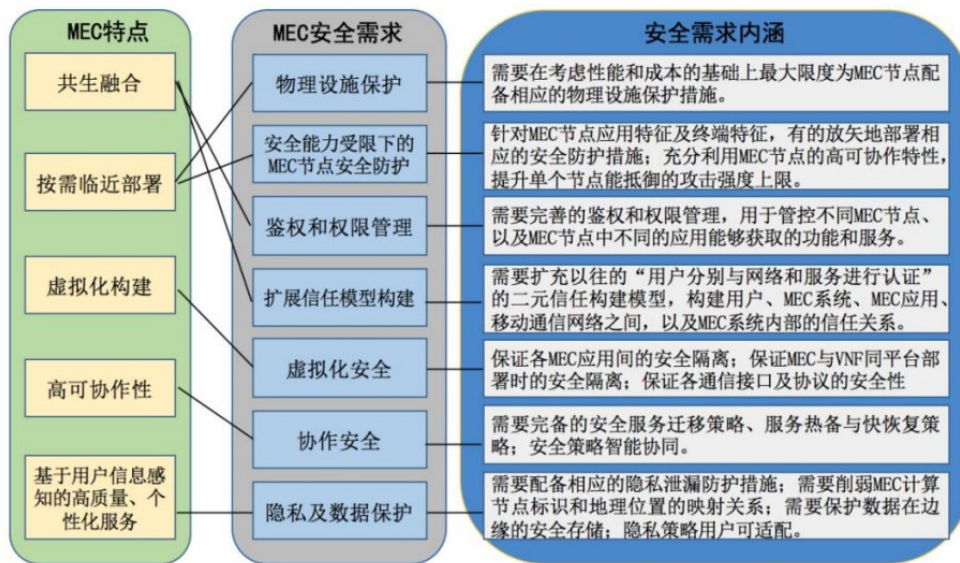
10 安全管理

10.1 概述

MEC 是通过将云计算能力和 IT 服务环境下沉到移动通信网络边缘，就近向用户提供服务，从而构建一个具备高性能、低延迟与高带宽的电信级服务环境。相对于运营商的网络，一般认为核心网机房处于相对封闭的环境，受运营商控制，安全性有一定保证。而接入网相对更易被用户接触，处于不安全的环境。MEC 更强调本地业务处理特性，使得数据在核心网之外终结，运营商的控制力减弱，攻击者可能通过边缘计算平台或的应用攻击核心网，造成敏感数据泄露、(D)DOS 攻击、拒绝服务等。所以，MEC 安全管理是必须要重点考虑的关键问题。

10.2 安全需求

5G MEC 的安全需求总结如图 4-10-1 所示^[1]。



5G MEC 特有的安全需求详细描述：

(1) 物理设施保护：MEC 按需临近部署的特点在为用户提供高质量服务的同时，也在客观上缩短了攻击者与 MEC 物理设施之间的距离，使得攻击者更有可能接触 MEC 物理设施，造成 MEC 物理设备损坏、服务中断、用户隐私 和数据泄露等严重后果。另一方面，广泛部署的 MEC 边缘计算节点同样面临着各种自然灾害（如台风、冰雹）和工业灾难的威胁。以上因素都可能直接破坏 MEC 硬件基础设施，造成服务的突然中断以及数据的意外丢失。因此需要在考虑性能和成本的基础上最大限度为 MEC 节点配备相应的物理设施保护措施。

(2) 安全能力受限下的 MEC 节点安全防护：由于性能、成本、部署灵活性要求等多种因素制约，单个 MEC 节点的安全防护能力（如可抵御的攻击种类，抵御单个攻击的强度等）受到限制。因此需要：针对 MEC 节点应用特征及终端特征（IoT 终端/移动智能终端），有的放矢地部署相应的安全防护措施；充分利用 MEC 节点的高可协作特性，通过例如基于智能协同的安全防护等技术，借助周边节点的空闲安全防护资源，提升单个节点能抵御的攻击强度上限。

(3) 扩展信任模型构建：MEC 系统与移动通信系统共生融合的部署方式扩充了以往的“用户分别跟网络和服务进行认证”的二元信任关系构建模型。需要构建用户、MEC 系统、MEC 应用、移动通信网络两两之间，以及 MEC 系统内部的信任关系。具体而言：需要构建 MEC 系统与 5G 网络间的信任关系以合法使用 5G 开放网络服务（如本地分流）向用户提供服务；需要构建 MEC 系统与 MEC 应用间的信任关系防止恶意应用接管用户服务；需要构建 MEC 系统与用户间的信任关系以确认 MEC 系统和用户的合法性。在 MEC 系统内部，需要构建 MEC 节点与 MEC 控制器间的信任关系，防止“伪 MEC 节点”恶意接入窃取用户和服务信息；需要构建 MEC 节点间的信任关系以支持节点间协作。

(4) 隐私及数据保护：MEC “基于用户信息感知的高质量、个性化服务”的特点在提供便利的同时也让 MEC 应用不可避免的接触到大量移动用户与设备的隐私和数据信息，如用户身份、位置、移动轨迹等。而对这些信息进一步挖掘后，还得到用户的作息规律、生活习惯、健康状况等诸多信息。因此，在 MEC 隐私及数据保护中，需要配备相应的隐私泄漏防护措施，

严控第三方 MEC 应用的行为，防止其泄漏、滥用用户的隐私及数据信息；需要通过动态身份标识和匿名等技术削弱 MEC 计算节点标识和地理位置的映射关系，防止第三方根据 MEC 节点位置推断用户的地理位置；需要确保数据在边缘的安全存储；需要向用户提供隐私及数据管理服务，确保隐私策略用户可适配。

10.3 安全框图

根据 ETSI 的 MEC 架构，参考如下图 4-10-2 所示，MEC 的安全威胁重点应考虑^[2]：

(1) 物理基础设施

物理基础设施安全方面，通过加锁、人员管理等保证物理环境安全。并对服务器的 I/O 进行访问控制，对物理网络进行安全组网规划、对物理主机进行安全隔离，及时更新操作系统安全漏洞并进行安全加固等。在条件允许时，可使用可信计算保证物理服务器的可信；

(2) 虚拟化基础设施

对 Host OS、虚拟化软件、Guest OS 进行安全加固，防止镜像被篡改，并提供虚拟网络隔离和数据安全机制。当虚拟机中部署容器时，还应考虑容器的安全，包括容器之间的隔离，容器使用 root 权限的限制等。

(3) MEC 平台

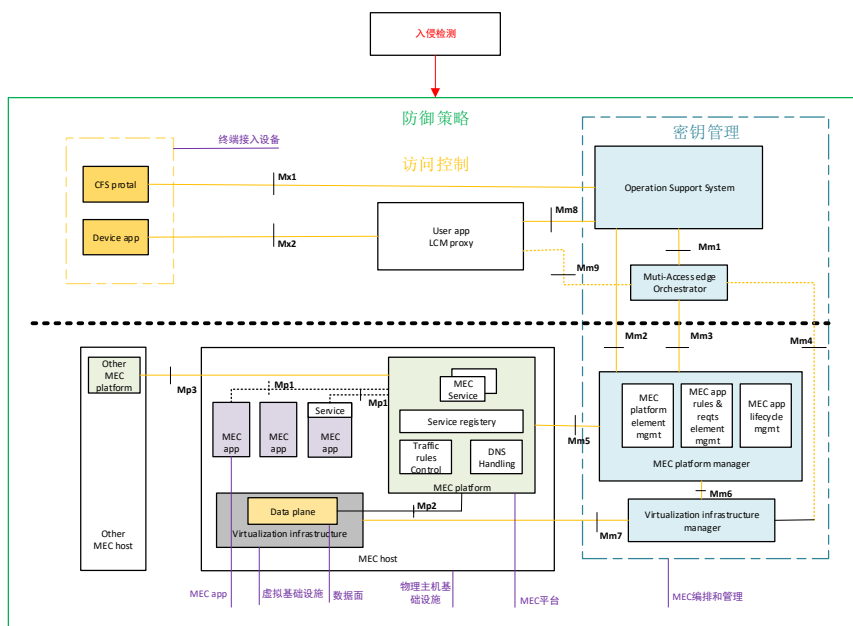
包含接口安全、API 调用安全、MEC 平台自身的安全加固以及敏感数据的安全保护、(D)DoS 防护，实现 MEC 平台与其他网元（如 ME a）间的通信数据的机密性、完整性、防重放，以及 MEC 平台的网络信息等敏感数据的安全保护、防(D)DoS 攻击等。

(4) MEC 编排和管理

包括接口安全、API 调用安全、数据安全和 MEC 编排和管理网元安全加固,实现对资源的安全编排和管理。与传统网络的安全管理一样，包含账号和口令的安全、授权、日志的安全等，保证只有授权的用户才能执行操作，所有操作记录日志。

(5) MEC 数据面

包含数据面网关的安全加固、接口安全、敏感数据保护以及物理接触攻击防护，实现用户数据能够按照分流策略进行正确的转发。



安全防护技术框架包括：入侵检测、访问控制、防御策略、密钥管理和协议与网络安全。

入侵检测主要用来监控和检测主机侧或网络侧的异常数据。在不同的网络环境下，入侵检测系统（intrusion detection system, IDS）的检测算法及应用部署有着不同的需求。

MEC 防御策略指的是保护 MEC 终端设备免遭来自外部的攻击，保护那些托管在 MEC 主机中的系统、应用程序，以及分布式部署在核心网或云数据中心的系统等，如身份认证与鉴权等。

MEC 访问控制指的是终端、MEC 主机、MEC 系统、MEC 数据和存储、以及编排管理等各组件模块之间的相互访问下的数据安全控制，确保 MEC 数据和隐私安全，如 RBAC 基于角色的访问控制、信任管理等。

密钥管理包括用于 MEC 网络通信过程的加密技术，如基于身份标识的密码技术、RSA 和 ECC 密码技术、基于配对的密码技术、格密码技术、多变量公钥密码技术等；用于 MEC 编排和管理相关的认证，如集群根证书（CA）。

协议与网络安全：MEC 系统网络架构包括终端用户支持多种网络接入技术（如 Wi-Fi、5G、4G、LoRa、有线等），以及 MEC 网络架构中的如 Internet、核心网等，MEC 整个网络系统都会受到外部和内部工具威胁，MEC 应用到的各种通信技术和协议都需要进行有效的安全防护。特别是在虚拟化环境中需要对不同租户之间进行网络隔离，针对不同租户的需求配置不同的网络安全策略。

10.4 关键技术

近年来提出的可能适用于 MEC 安全管理关键技术主要包括：数据安全、访问控制、身份认证和隐私保护^[3]、错误容忍和恢复等，如图 4-10-3 所示。

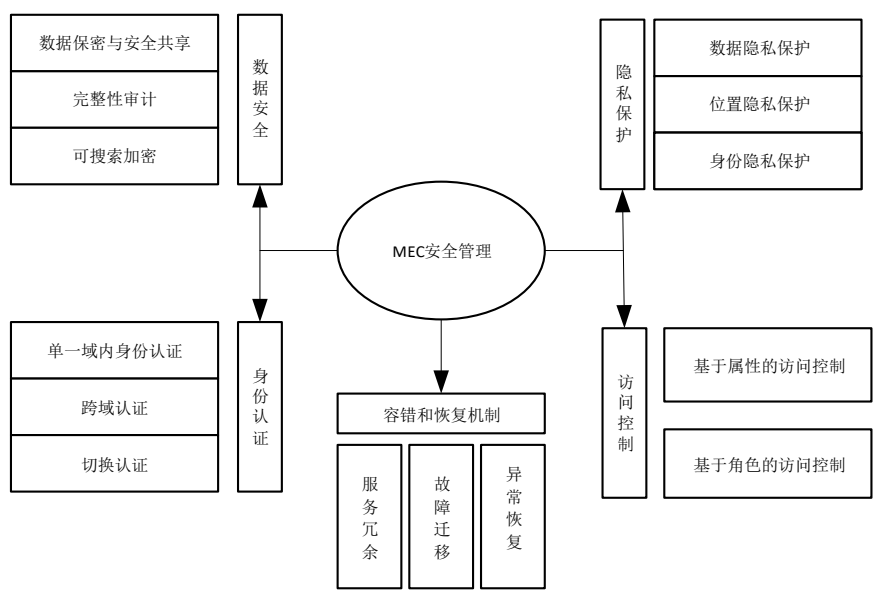


图 4-10-3 MEC 安全管理关键技术

1) 数据安全 (data security)

数据安全是创建安全边缘计算环境的基础，其根本目的在于保障数据的保密性和完整性。主要针对外包数据的所有权和控制权分离化、存储随机化等特性，用于解决数据丢失、数据泄露、非法数据操作等问题。同时，在此基础上允许用户进行安全数据操作。数据安全的主要内容包括数据保密性与安全共享、完整性审计和可搜索加密

现有的数据保密性和安全数据共享方案通常采用加密技术来实现，其常规流程是由数据所有者预先对外包数据进行加密处理和上传操作，并在需要时由数据使用者解密。常用的数据加密 (cryptography) 算法有基于属性加密 (ABE, attribute-based encryption)、代理重加密 (PRE, proxy re-encryption key) 和全同态加密 (FHE, fully homomorphic encryption) 算法等；

完整性审计是指当用户的数据存储到 MEC 主机或云数据中心之后，通过数据完整性审计确保外包存储数据的完整性和可用性。

可搜索加密 (SE, searchable encryption) 可以保障数据的私密性和可用性，并支持对密文数据的查询与检索。

安全排名搜索是指系统按照一定的相关度准则 (如关键字频率) 将搜索结果返回给用户。安全排名搜索提高了系统的适用性，符合边缘计算环境下的隐私数据保护的实际需求。

2) 身份认证 (identity authentication)

终端用户要想使用 MEC 所提供的服务，首先要进行身份认证。由于 MEC 是一种多信任域共存的分布式交互计算环境，因此，不仅需要为每一个实体分配身份，还需要考虑到不同信任域之间的相互认证。身份认证的主要研究内容包括单一域内身份认证、跨域认证和切换认证。

单一信任域内的身份认证主要用于解决每个实体的身份分配问题，各实体首先要通过授权中心的安全认证才能够获取存储和计算等服务。

跨域认证适用于不同信任域实体之间的认证机制研究还处于初级阶段，尚未形成较为完善的研究脉络和理论方法。

切换认证就是为了解决高移动性用户身份认证的一种认证移交技术，因此，对切换认证技术的研究能够为边缘计算中边缘设备的实时准确认证提供有力保障，同时在认证移交过程中的保护用户身份隐私。

3) 隐私保护 (privacy-preserving)

MEC 中的授权实体并不都是可信的，而用户的身份信息、位置信息和私密数据都存储在这些半可信 (honestbut-curious) 实体中，极易引发隐私问题。其主要内容包括数据隐私保护、位置隐私保护和身份隐私保护。

数据隐私保护指的是在保证用户隐私不被泄露的同时允许用户对数据进行各类操作（如审计、搜索和更新等）。

位置隐私保护指的是在用户终端的移动性，以及 MEC 提供的位置服务的同时能够不会暴露用户的真实位置。

身份隐私的保护实现了用户身份管理和隐私保护。

4) 访问控制 (access control)

访问控制是确保系统安全性和保护用户隐私的关键技术和方法。访问控制系统在原则上应适用于不同信任域之间的多实体访问权限控制，同时还应考虑地理位置和资源所有权等各种因素。当前比较热门的访问控制方案包括基于属性和基于角色的访问控制。

访问控制通常采用密码技术来实现，传统的密码技术并不适用于分布式并行计算环境，而属性加密 (ABE) 能够很好地适用于分布式架构，实现细粒度数据共享和访问控制。

基于角色的访问控制通过双重权限映射机制，即用户到角色和角色到数据对象上的权限映射来提供灵活的控制和管理。

5) 错误容忍与恢复

错误容忍与恢复机制指的是，在 MEC 系统被恶意攻击、错误的配置、代码健壮性不足、以及 MEC 应用软件异常时导致 MEC 服务中断时，能够通过 MEC 服务冗余配置、故障转移和异常恢复等机制实现连续性服务，甚至零中断服务。

10.5 研究进展

MEC 系统包括网络设施、服务设施、虚拟化设施及用户终端等多个系统层级。同时，MEC 系统的去中心化、分布式特征以及移动性支持、位置感知、服务注册和发现、本地分流、以及无线网络开放等特性。相对于传统云计算，MEC 的安全与隐私保护面临着更严峻的挑战。

刊文	内容概述
Security in mobile edge caching with reinforcement learning ^[4]	本文研究 MEC 系统中的攻击模型，重点研究移动卸载和缓存过程。在本文中，提出了应用强化学习 (RL) 技术的安全解决方案，以提供安全卸载到边缘节点对抗干扰攻击。还提出了轻量级认证和安全协作缓存方案来保护数据隐私。评估了基于 RL 的移动边缘缓存安全解决方案的性能，并讨论了未来需要解决的挑战。

刊文	内容概述
Cloud and MEC Security ^[5]	边缘云和 MEC 安全研究综述
刊文	内容概述
Security and privacy challenges in the internet of things [security and privacy matters] ^[6]	物联网的安全及隐私保护挑战研究综述
刊文	内容概述
Mobile edge computing, Fog et al.: A survey and analysis of security threats and challenges ^[7]	文章从整体角度分析了影响边缘计算的安全威胁和挑战，如雾计算，移动边缘计算和移动云计算
刊文	内容概述
The extended cloud: Review and analysis of mobile edge computing and fog from a security and resilience perspective ^[8]	探讨了与安全性和弹性相关的机制的方法，特别是基于异常检测和策略的弹性管理。
刊文	内容概述
Resource allocation and computation offloading with data security for mobile edge computing ^[9]	本文研究提出了一个具有数据安全性的多使用者资源分配与计算卸载模式。在此基础上，提出了一个综合考虑安全性、计算卸载和资源分配的综合模型，使整个系统的时间和能量消耗最小化。最后，提出了一种基于详细过程的卸载算法，以确定多用户系统的最优计算卸载决策。仿真结果表明，与本地执行和完全卸载方案相比，该模型和算法能显著提高整个系统的性能。
刊文	内容概述
Game theoretic path selection to support security in device-to-device communications ^[10]	一种用于 D2D 通信的新颖路由协议，该协议可以优化路由决策以显著改善恶意软件检测。该协议基于游戏理论模型，根据缓解恶意软件和花费在恶意软件检测上的能量来识别最佳网络路径。
刊文	内容概述
Security and privacy issues of fog computing: A survey ^[11]	雾计算的安全性和隐私问题调查研究综述
刊文	内容概述
Edge and Fog Computing in Critical Infrastructures: Analysis, Security Threats, and Research Challenges ^[12]	关键基础设施中的边缘和雾计算（Fog and Multi-Access Edge Computing, FMEC）：分析、安全威胁，以及研究挑战

10.6 参考文献

[1] 5G 信息安全白皮书[EB/OL]. [http://eversec.com.cn/wp-content/uploads/2019/03/L-](http://eversec.com.cn/wp-content/uploads/2019/03/L-5G%E4%BF%A1%E6%81%AF%E5%AE%89%E5%85%A8%E7%99%BD%E7%9A%AE%E4%B9%A6.pdf)

[5G%E4%BF%A1%E6%81%AF%E5%AE%89%E5%85%A8%E7%99%BD%E7%9A%AE%E4%B9%A6.pdf](http://eversec.com.cn/wp-content/uploads/2019/03/L-5G%E4%BF%A1%E6%81%AF%E5%AE%89%E5%85%A8%E7%99%BD%E7%9A%AE%E4%B9%A6.pdf)

[2] 庄小君, 杨波, 王旭, 等. 移动边缘计算安全研究[J]. 电信工程技术与标准化, 2018, 31(12): 38-43.

[3] 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.

[4] Xiao L, Wan X, Dai C, et al. Security in mobile edge caching with reinforcement learning[J]. IEEE

Wireless Communications, 2018, 25(3): 116-122.

-
- [5] Okwuibe J, Liyanage M, Ahmad I, et al. Cloud and MEC Security[J]. A Comprehensive Guide to 5G Security, 2018: 373.
- [6] Lee J H, Kim H. Security and privacy challenges in the internet of things [security and privacy matters][J]. IEEE Consumer Electronics Magazine, 2017, 6(3): 134-136.
- [7] Roman R, Lopez J, Mambo M. Mobile edge computing, fog et al.: A survey and analysis of security threats and challenges[J]. Future Generation Computer Systems, 2018, 78: 680-698.
- [8] Shirazi S N, Gougolidis A, Farshad A, et al. The extended cloud: Review and analysis of mobile edge computing and fog from a security and resilience perspective[J]. IEEE Journal on Selected Areas in Communications, 2017, 35(11): 2586-2595.
- [9] Elgendy I A, Zhang W, Tian Y C, et al. Resource allocation and computation offloading with data security for mobile edge computing[J]. Future Generation Computer Systems, 2019, 100: 531-541.
- [10] Panaousis E, Karapistoli E, Elsemary H, et al. Game theoretic path selection to support security in device-to-device communications[J]. Ad Hoc Networks, 2017, 56: 28-42.
- [11] Yi S, Qin Z, Li Q. Security and privacy issues of fog computing: A survey[C]//International conference on wireless algorithms, systems, and applications. Springer, Cham, 2015: 685-695.
- [12] Tedeschi P, Sciancalepore S. Edge and Fog Computing in Critical Infrastructures: Analysis, Security Threats, and Research Challenges[C]//2019 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW). IEEE, 2019: 1-10.