



# 面向 5G 的边缘计算及部署思考

## Thoughts on 5G Edge Computing and Deployment

马洪源/MA Hongyuan

(中国移动通信集团设计院有限公司, 北京 100080)  
(China Mobile Group Design Institute Co., Ltd., Beijing 100080, China)

DOI: 10.12142/ZTETJ.201903011

网络出版地址: <http://kns.cnki.net/kcms/detail/34.1228.TN.20190613.1731.002.html>

网络出版日期: 2019-06-13

收稿日期: 2019-04-26

**摘要:** 边缘计算对于运营商而言是一种网络架构和业务模式的创新。基于运营商的运维需求, 针对 5G 的边缘计算提出了一套系统化的解决方案。5G 边缘计算的部署基于业务需求和场景, 并结合网络需求、边缘基础设施、运营模式及维护管理需要, 是性能与投资的均衡考虑。强大的生态系统是 5G 边缘计算发展的保障, 完善的基础设施、灵活的网络和平台能力以及丰富的边缘应用是推动边缘生态繁荣的关键因素。

**关键词:** 边缘计算; 5G; 部署; 边缘生态

**Abstract:** Edge computing is an innovation in network architecture and business models for operators. A systematic solution based on the operation and maintenance requirements of operators is proposed in the 5G edge computing. Based on the business requirements and scenarios, the deployment of 5G edge computing combines with network requirements, edge infrastructure, operation mode and maintenance management needs. Meanwhile, the balance of performance and investment should also be considered. The powerful ecosystem is the guarantee of the development of the 5G edge, and the perfect infrastructure, the flexible network and the platform ability, and the rich edge application are the key factors to promote the edge ecological prosperity.

**Key words:** edge computing; 5G; deployment; edge computing ecology

随着网络的不断拓展, 越来越多的连接需要在边缘侧分析、处理与储存。随着终端能力的不断提升以及流量资费的进一步下调, 大流量业务将会对用户月均流量消费额(DOU)产生直接拉动效应; 预计在 5G 规模商用期, 单用户平均流量带宽将达到 4G 网络的 5~10 倍, 对回传网络产生巨大的压力。

当前的网络架构和移动技术对网络优化并不充分, 基站到核心网往往距离数百千米, 途径多重汇聚、转发设备, 再加上不可预知的拥塞和抖动, 难以保障一些对时延、可靠性要求较高的行业客户场景。

更多的业务机会, 更多的连接、

利用并赋能, 更好的业务体验需求均驱动了边缘计算的发展。而边缘计算作为一种网络架构和商业模式的创新, 是 5G 网络服务垂直行业的利器之一。边缘计算是运营商提升网络价值的一次契机, 将有可能推动产业链价值的重分配。

### 1 面向 5G 的边缘计算技术

边缘计算是一种在物理上靠近数据源头进行数据处理的方法, 是一种分布式计算架构。它能够进一步减小传输时延, 提高网络运营效率, 提高业务分发/传送能力, 优化/改善终端用户体验, 满足行业用户在数字化变革过程中对业务实时、智能、数据聚合与互操作、安全与隐

私保护等方面的关键需求。这与 5G 网络拓展垂直行业、面向服务的理念高度吻合, 边缘计算作为 5G 原生功能将有助于实现应用本地化、内容分布化和计算边缘化。

#### 1.1 面向 5G 的边缘计算为运营商提供了一个系统化解方案

边缘计算相关标准在 4G 网络部署初期并未被考虑, 欧洲电信标准化协会(ETSI)把移动网络和边缘计算的融合留给厂商实现也导致边缘计算方案各有不同。4G 边缘计算以最小化网络影响为设计出发点, 方案具有局限性。5G 边缘计算由 5G 核心网(5GC)、边缘计算平台和用户设备(UE)协同实现, 满足边

缘场景下的计费、合法侦听、移动性管理和服务质量(QoS)需求。

为了更好地发挥边缘计算系统的能力,5G 边缘计算借鉴了 ETSI 对边缘计算的参考架构和成熟理念,让用户平面功能(UPF)作为数据面锚点成为 ETSI 与第3代合作伙伴计划(3GPP)融合的关键点。在 5G 边缘计算中 3GPP 定义了网络架构,支持数据路由与能力开放;ETSI 定义了平台架构,边缘计算平台实现数据网络(DN)和应用功能(AF)。

边缘计算功能在初期的 5G 网络中没有专属的网络功能(NF),而是分散在接入和移动性管理功能(AMF)、会话管理功能(SMF)、策略控制功能(PCF)、网络业务呈现功能(NEF)、AF、UPF 的功能中。UPF 的选择与控制通过边缘计算平台与 5G 核心网(5GC)交互完成,边缘计算业务受集中部署的 5GC 统一管理和控制。边缘计算平台作为 5G 网络中的一个 AF,通过 NEF/PCF->SMF->UPF 路径管理 PDU 会话、控制策略设置,部署在边缘计算平台上的边缘应用可以属于一个或多个网络切片<sup>[1,2]</sup>。

### 1.2 5G 为边缘计算提供了灵活可控的网络能力

5G 网络采用 C/U 分离架构,用户面通过按需分布式部署,实现流量的本地卸载,从而支持端到端毫秒级时延。

5G 边缘计算为本地流量卸载提供了 3 种方案:上行分流、IPv6 多归属以及本地数据网。

(1)上行分流根据 SMF 下发的过滤规则,通过检查数据包目的 IP 地址进行分流。上行分流方案中 SMF 决定在 PDU 会话中插入上行分类器(UL-CL);UPF 执行具体流程,支持 UL-CL 功能的 UPF 按照 SMF 提供流量模板匹配业务流;UL-CL 功能提供到不同 PDU 锚点的业务前传和下行流量汇聚。该方案中 1 个 PDU 会话仅 1 个 IP 地址(不考虑双栈),终端地址可以不变,减少应用层的交互。该方案适用于访问本地业务场景,如本地内容访问企业网、增强移动宽带(eMBB)场景本地分流业务、车联网等,通常采用隧道方式。(2)IPv6 多归属方案中分支点(BP)根据 SMF 下发的过滤规则,通过检查数据包源 IP 地址进行分流。IPv6 多归属主要解决 1 个 PDU 会话、多个 IPv6 地址的问题。该方案适用于物联网、高可靠性专网等场景;但由于要采用 IPv6,目前实施难度较大。(3)本地数据网分流通过终端判断自身位置,如终端处在本地数据网络(LADN)服务范围,则发起携带 LADN 数据网络名称(DNN)的会话建立请求。

3 种分流方式都需要 UPF 和 5GC 的参与,但对终端和网络要求依次升高。

与此同时,5G 网络为满足按需的业务连续性,定义了 3 种会话与业务连续性(SSC)模式,可以针对不同的边缘计算业务保持 PDU 会话连接“锚点一直不断”“先拆后建”或者“先建后拆”。针对车到万物(V2X),如采用边缘计算技术,终端移动会导致 UPF 的频繁切换,此

时 SSC mode3 模式可以在网络域保障业务连接不断。

### 1.3 能力开放和边缘智能是 5G 边缘计算的价值链延伸

5G 边缘计算能力开放包括网络能力开放、信息技术(IT)资源开放以及管理开放。

网络能力开放主要指网络及用户信息开放、业务及资源控制功能开放。例如,通过将移动用户的位置信息开放,将人、物和数据之间的连接情景化,进一步拓展基于位置的精准营销服务及室内定位业务;通过将无线小区的负载信息、链路质量的实时及统计信息、网络吞吐量的实时及统计信息开放开始实现业务及网络质量优化。

5G 边缘计算的开放通过网络能力开放通过平台中间件获取底层网络信息,统一发送给 NEF 实现能力开放调度与分发,NEF 再通过北向标准的 Restful 应用程序编程接口(API)开放给第三方边缘应用。边缘应用将处理后的数据信息反馈给 NEF,NEF 基于服务化接口通过 5GC 下发给边缘计算平台及 UPF 执行相关策略。

资源开放包括 IT 基础资源的开放、管理。边缘应用通过边缘计算的云管理系统实现底层 IT 资源规划部署、动态优化和业务编排。另外,考虑到 5G 边缘计算生态多样性,边缘计算还要提供图形处理器(GPU)、现场可编程门阵列(FPGA)等不同的硬件资源,同时还需叠加多种软、硬件加速技术,提供端到端的加速能力,以便更好地支撑资源

能力开放。

管理开放指 5G 边缘计算管理系统通过对控制模块进行路由策略设置,针对不同用户、设备或者第三方的应用需求,实现对移动网络数据平面的控制。管理开放系统包括边缘计算实例生命周期管理、边缘计算平台中间件的创建、消亡以及第三方调用授权。

边缘能力开放需要实现边缘计算平台 API 的丰富和统一:丰富指需要业务丰富多样,既能通过网络能力 API 提供网络能力,又能通过行业 API 提供应用辅助能力,有力地促进应用的创新;统一指接口标准统一,便于调用、易扩展、跨平台,实现规模化效益。

5G 网络是一种技术的收敛、应用的发散。当边缘应用场景海量参数无法依靠人工来优化的时候,则必须依靠边缘智能来解决。基于 5G、借助人工智能技术的边缘智能是边缘计算的更高价值体现:通过在 5G 网络边缘部署具有人工智能芯片的人工智能型边缘计算平台,提供面向 5G 本地业务应用的人工智能运算和分析能力,引入边缘智能的 5G 边缘计算将有助于提升运营商网络智能化和网络自动化,进一步地延伸网络价值链。当边缘计算成为边缘智能,可以使得边缘系统具备自治自律的能力、自给自足的算力以及智能,并将使得边缘应用摆脱“云计算”而相对独立地进行运营。

## 2 5G 边缘计算部署关键问题

边缘计算是个复杂的生态系

统。5G 边缘计算的部署以应用为导向,受时延、带宽、数据安全及边缘基础设施等因素影响,是满足业务指标、兼顾投资效益和运维需求的均衡考虑。

### 2.1 业务需求和场景选择是首要考虑因素

eMBB 场景中的边缘计算更多考虑个人用户,针对大视频类业务、流量密集区域,网络和内容协同下沉。随着 5G 标准进一步完善,以及垂直行业的拓展,5G 边缘计算开始在行业用户发力,但对于行业和场景的选择至关重要。无论是作为物联网边缘网关,还是与网络切片结合提供可定制的专网服务,5G 边缘计算部署都需要充分考虑该业务场景的市场容量及网络部署可行性。通过与特定行业的合作,加深对行业的理解,从行业边缘业务向核心业务渗透,通过优势互补提升合作空间。目前智能制造、智慧城市、直播游戏和车联网 4 个垂直领域对边缘计算的需求最为明确<sup>[9]</sup>。

### 2.2 网络指标与投资效益需要均衡考虑

时延和带宽是 5G 边缘计算初期较为受关注的 2 个主要网络指标。5G 网络通过引入新的空口和传输网络技术,能够直接减少网络传输时延。5G 边缘计算通过部署在接入网和核心网之间不同位置,可以不同程度地减少网络中间层级,简化网络架构,能够满足低时延的业务要求,同时减少互动性需求对网络时延累加效应的影响,增加

网络确定性。5G 边缘计算是网络与内容的下沉,但减少回传带宽不仅仅是因为网络结构的调整,还需要网络与内容的协同。通过加大本地内容源引入、优化卸载算法,可以有效提升边缘节点对内容分发网络(CDN)命中率,从而提高数据流量本地分流比例。

有限的自主性能能够保证边缘计算系统相对的独立,满足行业用户对边缘应用的部分管理需求。安全和隐私保护也是边缘计算部署中需要考虑的重要因素,边缘场景中很多数据是企业保密信息、个人隐私或受到监管的信息。通过网络流量隔离、部署安全防护设备以及增加安全管理手段可以满足边缘场景对网络安全的诉求。

但 5G 边缘计算的部署并非是一味地满足网络指标,而是需要实现网络性能与建设成本的平衡;没有经济效益的技术是难以持久的。因此 5G 边缘计算并非越靠近网络边缘越好,有效降低用户网络使用成本也是 5G 边缘计算部署需要考虑的重要因素。一方面,边缘计算越靠近网络接入侧,中间环节就减少,从而用户体验就越好;另一方面,越靠近网络边缘,同时接入的用户就会变少,网络边际效益降低,网络管理复杂,总体成本上升。当数据仅在本地有价值时,边缘计算能够更近距离地处理和储存原始数据,从而节约成本。当需要处理海量数据时,5G 边缘计算通过本地分析和过滤能够减少送往数据中心的数据量,这不仅降低了组网成本,而且为更重要的流量处理保留了有限



的网络带宽。

### 2.3 边缘基础设施是 5G 边缘

#### 计算部署的前提和重要保障

边缘基础设施是 5G 边缘计算部署的重要载体,包括边缘机房、边缘配套网络、边缘计算服务器等。

相较于运营商集中化的核心机房,边缘机房在温度、承重、可用性和供电等诸多方面均存在较大差异。根据边缘计算在运营商网络中部署的位置,边缘机房主要包括地市核心机房、重要汇聚机房、普通汇聚机房和站点机房。其中前 2 类机房条件相对较好,5G 边缘计算可以考虑使用通用硬件服务器,部署完整的组网及管理设备,实现不同流量物理隔离。机房单机柜则建议按照 3~5 kW 考虑空调散热设计,并优选使用高功率密度电源设备以便节省占地面积。对于机房条件受限和改造相对困难的普通传输机房和站点接入机房,可以考虑定制化的高密度组件硬件服务器,例如多核中央处理器(CPU)( $\geq 16$ 核)、大容量内存以及大容量固态存储器(如 SSD、Flash)。同时,边缘计算服务器应具备低功耗( $< 500$  W),可以支持更宽的温度范围,支持无风扇散热能力,并能够降低散热要求。目前行业联合推出的开放电信 IT 基础设施(OTII)服务器方案就是一种积极的探索尝试。

5G 边缘计算需要网络支持三层路由,UPF 需要通过 IP 专网接受 5GC 的统一管理。当部署在区县一级或更接近边缘侧时,UPF 和边缘平台近期可通过切片分组网

(SPN)/分组传送网(PTN)或建立专线方式接入 IP 专网,远期需要考虑承载网边界的下移。

5G 边缘计算硬件服务器可以采用以数据存储为主的存储型平台、以数据处理为主的计算型平台,以及兼顾数据处理和数据存储的综合平台。针对不同的业务场景,服务器也可以采用不同的处理架构,Intel 和 ARM 分别侧重网络侧及现场级边缘计算:X86 设备侧重网络侧边缘云,嵌入式系统侧重现场级边缘智能。边缘存储有本地存储、磁盘阵列和分布式存储(含 VIM 内置)多种方案。边缘存储方案根据业务场景确定存储服务器类型和架构,并结合边缘机房空间、温度、承重等条件,进行功耗、尺寸、硬盘选择、接口类型等选配选择与制定。同时,根据业务特点对部分功能、性能、可靠性进行增强。为避免对单一架构的过度依赖,边缘服务器应鼓励基于不同处理架构的自研设备。但无论采用何种平台和架构,5G 边缘计算基础设施都应该以开放、标准的云计算平台为基础,实现异构网络共平台部署,快速简单地第三方应用搭建业务环境。

### 2.4 运营模式和管理需要在 5G 边缘计算部署过程中逐步完善

从云计算角度理解 5G 边缘计算的运营,存在基础设施即服务(IaaS)、平台即服务(PaaS)和软件即服务(SaaS)等不同模式。运营商可以提供本地分流服务、边缘机房出租、边缘硬件基础设施、统一 IaaS 能力、PaaS 能力等不同的服务。运

营商可以通过建设基础 IaaS 能力+通用 PaaS 平台+第三方行业 PaaS 能力的方式,对不同企业客户提供不同商业模式。例如,针对大型企业可以提供网络能力+基础 IaaS 能力方式按需建设;针对中小企业和自有业务可以通过边缘网络+IaaS+PaaS 平台的方式统一部署和规划。

边缘节点存在电信业务、IT 类应用、边缘支撑及管理等多类业务,对应的运营实体可能不一致。而组织架构会影响技术架构,因此运营组织架构的调整需要在 5G 边缘计算的部署中优先考虑,并在后续运维中不断优化。

5G 边缘计算的管理可以采用“集中+分布”的方式部署:在地市级边缘节点统一部署边缘节点管理平台,负责业务分发、部署、升级、统一门户和资源视图等操作。考虑到区县以下边缘节点资源有限,无论是业务管理还是资源管理,开销都不宜过大。IaaS 可以采用轻量化 OpenStack、软件定义网络(SDN)和容器技术;PaaS 仅提供基本的运维环境和运维工具,支持协议转换、异构网络接入、数据采集等功能。同时,还需要考虑边缘节点无人值守情况下的远程维护。

5G 边缘计算存在混合云部署的情况,而目前网络功能虚拟化管理和编排(MANO)更多地聚焦在电信网络;因此对于部署在边缘的第三方 PaaS 平台及第三方应用的管理仍需要进一步探索和完善。

## 3 结束语

边缘网络已成为信息产业争夺

的一个重要领域。面向 5G 的边缘计算提供无线通信网络与互联网的融合应用平台,开拓新的业务领域,增加新的收入。可靠的 5G 网络接入能力、灵活可控的网络和平台能力、分布广泛的边缘基础设施是运营商切入边缘计算的重要抓手。通过规模优势降低使用成本,运营商能够以专业的网络运维和管理能力、更加开放的平台能力在边缘场景提供经济高效的网络服务。

面向 5G 的边缘计算,运营商天然具备网络能力,并体现了连接的价值。运营商需要积极拓展平台能力,实现“大中台”的部署,不断提升

控制的价值。未来,运营商还需要引领行业能力,推动生态繁荣和云网协同,发挥融合和开放的价值。

针对 5G 边缘基础设施规划,建议采用点面结合,针对地市核心/骨干汇聚机房,“全面”启动基础设施储备,并基于通用硬件平台和用户面下沉,发挥网络+内容的协同优势,快速构建边缘计算能力;针对站“点”机房,则按需进行资源规划,深入理解行业,拓展 5G 边缘新兴业务,建立新的产业链和生态圈。

#### 参考文献

- [1] 3GPP. Technical Specification Group Services and System Aspects; System Architecture for the 5G System; Stage 2 (Release 15):V15.1.0:

3GPP TS 23.501 [S]. 2018

- [2] 3GPP. Technical Specification Group Services and System Aspects; Procedures for the 5G System; Stage 2 (Release 15):V15.1.0: 3GPP TS 23.502 [S]. 2018  
 [3] 中国移动边缘计算技术白皮书[R]. 中国移动边缘计算开放实验室, 2019  
 [4] ETSI. Mobile Edge Computing (MEC) Terminology[EB/OL].(2016-12-23)(2019-04-20). [http://www.etsi.org/deliver/etsi\\_gs/MEC/001\\_099/001/01.01.01\\_60/gs\\_MEC001v010101p.pdf](http://www.etsi.org/deliver/etsi_gs/MEC/001_099/001/01.01.01_60/gs_MEC001v010101p.pdf)

#### 作者简介



马洪源,中国移动通信集团设计院有限公司网络所高级咨询设计师;主要从事网络咨询、规划和工程设计工作,在移动通信核心网、NFV、IMS、5G、边缘计算等领域积累了丰富的咨询设计和研究经验;发表论文 10 余篇。

#### ← 上接第 67 页

- mcom.2016.7565185  
 [29] ABANI N, BRAUN T, GERLA M. Proactive Caching with Mobility Prediction under Uncertainty in Information-Centric Networks [C]//Proceedings of the 4th ACM Conference on Information-Centric Networking. USA: ACM, 2017: 88-97  
 [30] POULARAKIS K, TASSIULAS L. Code, Cache and Deliver on the Move: A Novel Caching Paradigm in Hyper-Dense Small-Cell Networks [J]. IEEE Transactions on Mobile Computing, 2017, 16(3): 675-687. DOI: 10.1109/tmc.2016.2575837  
 [31] LI X, WANG X, LI K, et al. Collaborative Hierarchical Caching for Traffic Offloading in Heterogeneous Networks[C]// Communications (ICC), 2017 IEEE International Conference on. USA: IEEE, 2017: 1-6  
 [32] JI M, CAIRE G, MOLISH A F. Wireless Device-to-Device Caching Networks: Basic Principles and System Performance[J]. IEEE Journal on Selected Areas in Communications, 2016, 1(34): 176-189. DOI: 10.1109/JSAC.2015.2452672  
 [33] JIANG J, ZHANG S, LI B, LI B. Maximized Cellular Traffic Offloading via Device-to-Device Content Sharing. IEEE Journal on Selected Areas in Communications, 2015, 34(1): 82-91. DOI:10.1109/JSAC.2015.2452493  
 [34] BASTUG E, BENNIS M. Living on the Edge:

The Role of Proactive Caching in 5G Wireless Networks[J]. IEEE Communications Magazine, 2014, 8(52): 82-89

- [35] HULL B, BYCHKOVSKY V, ZHANG Y, et al. CarTel: A Distributed Mobile Sensor Computing System[C]//Proceedings of the 4th International Conference on Embedded Networked Sensor Systems. USA: ACM, 2006: 125-138  
 [36] ZHANG Q, ZHAO J H. A Model for Automatic Collection and Dynamic Transmission of Traffic Information based on VANET[C]// Intelligent Transportation Systems (ITSC), 2012 15th International IEEE Conference on. USA: IEEE, 2012: 373-378. DOI: 10.1109/ITSC.2012.6338711  
 [37] FERNANDO T, MERCEDES V, CRISTINA S, et al. A Cooperative Approach to Traffic Congestion Detection with Complex Event Processing and VANET. IEEE Transactions on Intelligent Transportation Systems, 2012, 13(2): 914-929. DOI:10.1109/TITS.2012.2186127  
 [38] REZAEI F, NAIK K, NAYAK A, et al. Effective Warning Data Dissemination Scheme in Vehicular Networks for Intelligent Transportation System Applications[C]//16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013). Netherlands: IEEE, 2013: 1071-1076. DOI: 10.1109/ITSC.2013.6728374  
 [39] DIKAIKOS M D, FLORIDES, A, NADEEM T, IFTODE L. Location-Aware Services over Vehicular Ad-Hoc Networks using Car-to-

Car Communication[J]. IEEE Journal on Selected Areas in Communications, 2007, 25(8): 1590-1602

- [40] KNORR F, BASELT D, SCHRECKENBERG M, et al. Reducing Traffic Jams Via VANETs[J]. IEEE Transactions on Vehicular Technology, 2012, 61(8): 3490-3498. DOI:10.1109/tvt.2012.2209690

#### 作者简介



唐洁,华南理工大学计算机科学与工程学院副教授,IEEE 无人驾驶学术委员会秘书长;目前研究方向为无人驾驶中的边缘计算和大数据计算系统;曾参与国家自然科学基金、广东省科技重大专项等课题;已发表论文 30 余篇。



刘少山,深圳普思英赛科技有限公司创始人兼 CEO、IEEE 无人驾驶学术委员会常务副主席、IEEE 高级会员等;目前研究方向为无人驾驶与机器人系统;已编写教材 2 本,发表论文 20 余篇,拥有专利 100 余项。