



专题：数据网络协议架构创新——NewIP

基于云、网、边融合的边缘计算新方案：算力网络

雷波, 刘增义, 王旭亮, 杨明川, 陈运清

(中国电信股份有限公司北京研究院, 北京 102209)

摘要: 边缘计算已经成为 5G 时代重要的创新型业务模式, 尤其是其低时延特性, 被认为是传统方案所不具备的, 因此边缘计算能够提供更多的服务能力且具有更为广泛的应用场景。但边缘计算与处于中心位置的云计算之间的算力协同成为新的技术难题, 即需要在边缘计算、云计算以及网络之间实现云网协同、云边协同, 甚至边边协同, 才能实现资源利用的最优化。在研究边缘计算算力分配和调度需求的基础上, 提出了基于云、网、边深度融合的算力网络方案, 并针对 AI 类应用给出了一个典型实施系统, 该方案能够有效应对未来业务对计算、存储、网络甚至算法资源的多级部署以及在各级节点之间的灵活调度。

关键词: 边缘计算; 云计算; 云网融合; 算力网络

中图分类号: TP393

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2019209

Computing network: a new multi-access edge computing

LEI Bo, LIU Zengyi, WANG Xuliang, YANG Mingchuan, CHEN Yunqing

Beijing Research Institute of China Telecom Co., Ltd., Beijing 102209, China

Abstract: Edge computing has become an important innovative business model in the 5G era, especially its low latency characteristics, which are considered to be unavailable in traditional solutions. Therefore, edge computing can provide more service capabilities and more application scenarios. However, the synergy of computing resources between edge computing and cloud computing has become a new technical problem, so it is necessary to realize cloud network collaboration, cloud edge collaboration and even edge collaboration between edge computing, cloud computing and network, so as to achieve the optimization of resource utilization. A computing network solution based on cloud, network and edge depth fusion was introduced, and a typical system for AI application was proposed, which could effectively cope with the future.

Key words: edge computing, cloud computing, cloud network convergence, computing network

1 引言

随着 5G 时代的到来, 业界在畅想 5G 将改变

社会时, 经常会提及边缘计算, 并将其视为改变通信信息服务模式的关键创新之一。据 IDC 发布的《数据时代 2025》报告预测, 到 2025 年 50%的数据将

收稿日期: 2019-07-15; 修回日期: 2019-08-20

基金项目: 国家重点研发计划基金资助项目 (No. 2018YFB1800100)

Foundation Item: The National Key Research and Development Program of China (No. 2018YFB1800100)

2019209-1

在网络边缘侧分析、处理与存储，与此同时边缘计算也被认为是 5G 与工业互联网、物联网等的重要结合点，能够推动相关产业带来飞跃性发展。但随着研究和实践的深入，边缘计算的概念已经不限于 5G 领域，扩展到了专线、PON、Wi-Fi、4G 等，即多接入边缘计算（multi-access edge computing）。

但目前业界对边缘计算的定义与内涵并没有形成一致意见，各标准组织或企业分别从自己的角度提出了不同的认识，如 ISO 认为边缘计算是一种将主要业务处理和数据存储放在网络边缘节点的分布式计算形式，ETSI 定义边缘计算是在靠近数据源或用户的地方提供计算、存储等基础设施，并为边缘应用提供云服务和 IT 环境服务，而国内的边缘计算产业联盟（ECC）则定义边缘计算是靠近物或数据源头的网络边缘侧，融合网络、计算、存储、应用核心能力的开发平台等。虽然大家都认可边缘计算是在网络边缘上提供计算服务这个基本观点，但对什么是网络边缘、边缘计算的功能包含什么、如何部署和实现都存在较大争议。邬贺铨院士在 2018 年年底提出了“十问”边缘计算，系统化地总结了边缘计算发展过程中遇到的多个关键重大问题。

在“十问”中，邬院士连续追问了两个关于计算能力部署和调度的问题，即第二问“计算能力是一级设置还是多级设置？”和第三问“计算能力如何在边缘计算和云计算之间优化配置？”。这里直指了一个边缘计算部署过程中最常见的问题，即算力的分配与调度问题。没有人认为边缘计算会完全替代云计算，因此自然存在这个疑问。笔者在推进边缘计算现场试点过程中，被一线生产运营部门问得最多的也是这个问题。

为了解决这个问题，一种方案是采用“云边协同”的方案，即将边缘计算同样划分为 IaaS（基础设施即服务）、PaaS（平台即服务）、SaaS（软件即服务）等多层，然后将 EC-IaaS（边缘计算的

基础设施即服务）与云端 IaaS 对接实现对网络、虚拟化资源、安全等的资源协同；EC-PaaS（边缘计算的平台即服务）与云端 PaaS 对接实现数据协同、智能协同、应用管理协同、业务管理协同；EC-SaaS（边缘计算的软件即服务）与云端 SaaS 对接实现服务协同。但在实践过程中，发现这种方案需要边缘计算节点具备复杂的云计算管理平台，但边缘计算所在的边缘机房一般环境受限，能够容纳的服务器资源有限，此方案需要将过多的资源用于管理和协同，其建设与维护成本可观，存在一定的局限性。

因此本文提出了一种新的解决思路，即利用云网融合技术以及 SDN/NFV 等新型网络技术，将边缘计算节点、云计算节点以及含广域网在内的各类网络资源深度融合在一起，减少边缘计算节点的管控复杂度，并通过集中控制或者分布式调度方法与云计算节点的计算和存储资源、广域网的网络资源进行协同，组成新一代信息基础设施，为客户提供包含计算、存储和连接的整体算力服务，并根据业务特性提供灵活、可调度的按需服务。采用这种方案构建新型信息基础设施架构，被称为“算力网络”，其能够根据客户需求，在云、网、边之间按需分配和灵活调度计算资源、存储资源以及网络资源。

2 边缘计算与算力调度需求

边缘计算并不是简单地将服务器放到边缘机房即可，目前普遍认为边缘计算应该具备三大关键指标，即“低时延、大带宽和低成本”，只有这样的方案，才能让客户和平台运营方双赢。因此针对这三大指标逐一进行分析。

首先，针对低时延指标。究竟时延需要多低，各方给出的建议数值不一致，有 2 ms、4 ms、6 ms 或者 10 ms 等。但业界主流认为至少客户流量不需要在广域网上绕行，即无论是 4G/5G 的移动接入方式，还是以 PON 为主的光接入方式，都希望



客户流量能够直接从接入位置就近送入边缘计算节点，无须到运营商原有的部署在汇聚或核心机房的业务接入控制网关处绕行。以光接入为例，理想的接入路径是流量直接从客户站点到 OLT 再到边缘计算节点，但现有的实践路径是从客户站点到 OLT 后，先上行到 MSE（综合业务网关，也可以是 SR（业务路由器），然后再回到 OLT 进入边缘计算节点，如图 1 所示。因此在边缘计算提出之时，就考虑采用 5GUPF 下移的方案来解决相关问题，同样采用光接入方式也需要 vBRAS、vCPE、vSR 分布式部署等解决方案。

其次，针对大带宽指标，随着 5G 网络建设的开展，移动接入和固网接入的双吉比特已经成为下一阶段网络服务的标准配置，因此接入带宽并不是困难所在。

最后，对于低成本指标，这里所指低成本不

是说边缘计算能够实现无条件的低成本。事实上，单以建设和运营来看，传统的网络边缘机房改造困难很多，比如电力引入、空调改造、承重加固等都存在很多问题，通常需要投入大量的资金进行改造，使得边缘计算节点在单位建设成本上要远高于集中建设的云计算节点。但另一方面，考虑在相同服务质量（如低时延和低抖动）要求下，传统云计算方案需要网络提供高品质的专线传输（如 OTN 等传输专线），才能按性能指标要求将业务流量送达集中部署的云计算节点，这样算来，高品质传输专线成本与云计算节点成本之和可能高于边缘计算成本，因此有部分观点从这个角度得出了边缘计算成本更低的结论。

综上所述，边缘计算能够提供低时延、大带宽的高品质服务，但由于它的单位算力的成本高于云计算的算力成本，因此在不计入网络连接产

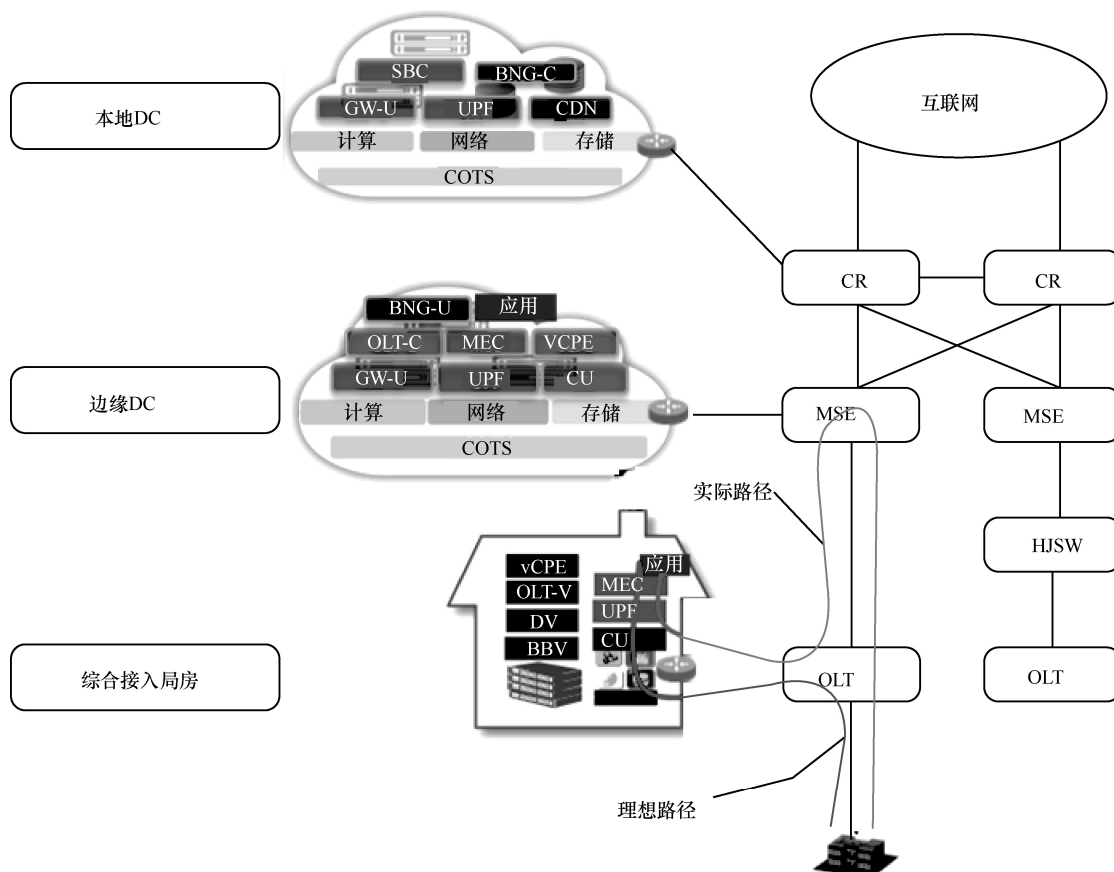


图 1 基于传统网络架构的边缘计算路径迁回示例

品价格时,边缘计算产品的定价应该高于云计算产品,所以边缘计算并不适用所有客户,更适用于那些愿意以一定费用来换取其业务所需的大带宽和低时延等高品质服务的客户。因此从业务角度来看,客户有合理安排其应用部署位置的需求,甚至采用混合部署模式,比如将非实时、不重要的运算放在云计算节点,将实时、重要的运算放在边缘计算节点。

以一个典型的AI应用为例,如图2所示,边缘计算节点负责数据的实时采集、计算和处理以及AI推理计算;而位置较远的云计算节点则负责大数据分析挖掘、数据共享,同时进行AI算法模型的训练和迭代以及用户个性化功能塑造等非实时工作;并且,云计算节点将迭代升级后的算法模型推送到边缘计算节点,使边缘计算节点更新和升级,完成自主学习闭环。

3 基于云、网、边深度融合的算力网络

从第2节分析中可以看出,多级算力部署是边缘计算乃至云计算发展的必然选择,因此在多级算力之间进行合理的算力分配与灵活调度,也就成为了边缘计算实施与部署过程中必不可少的一环。但现有的云计算服务体系中,尤其IaaS层面,还停留在让客户自行选择应用部署位置的阶段,没有提供按需分配和调度算力的手段。究其原因,因为现有的云计算体系并未将广域网(如电信运营商的接入网、城域网、骨干网)纳入整体的管控中,而各方所提的云网融合还处在一个相对简单的初级阶段,通常需要建设一个横跨云

网和网管的超级协同编排系统,运营难度相对复杂。因此有必要从底层架构开始,重新考虑和设计云、网、边深度融合方案,用以实现算力等基础信息资源的分配与调度,构建“算力网络”,成为一种新的技术发展方向。

目前,业界尚无对算力网络的标准定义,但本文认为算力网络需要满足以下4个特征要求。

(1) 资源抽象

算力网络需要将计算资源、存储资源、网络资源(尤其是广域范围内的连接资源)以及算法资源等都抽象出来,作为产品的组成部分提供给客户。

(2) 业务保证

以业务需求划分服务等级,而不是简单地以地域划分,向客户承诺诸如网络性能、算力大小等业务SLA,屏蔽底层的差异性(如异构计算、不同类型的网络连接等)。

(3) 统一管控

统一管控云计算节点、边缘计算节点、网络资源(含计算节点内部网络和广域网络)等,根据业务需求对算力资源以及相应的网络资源、存储资源等进行统一调度。

(4) 弹性调度

实时监测业务流量,动态调整算力资源,完成各类任务高效处理和整合输出,并在满足业务需求的前提下实现资源的弹性伸缩,优化算力分配。

总结起来,算力网络是“一种根据业务需求在云、网、边之间按需分配和灵活调度计算资源、

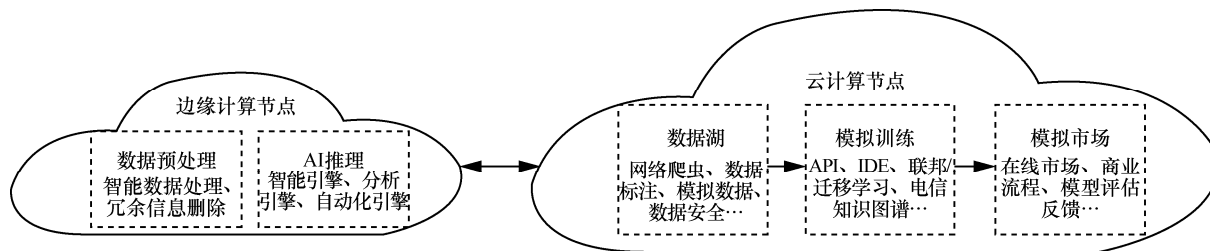


图2 AI应用在边缘计算与云计算节点之间的混合部署方案示例



存储资源以及网络资源的新型信息基础设施”。典型的算力网络示意图如图 3 所示。

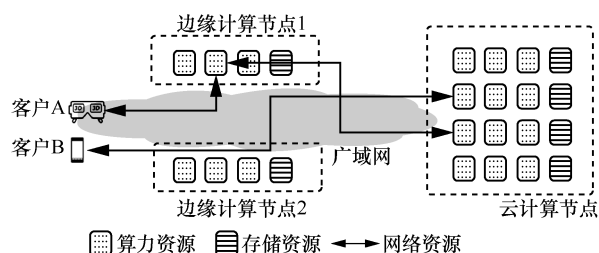


图3 算力网络示例

在图 3 中，客户 A 需要低时延、大带宽的 VR 应用，因此算力网络在客户 A 的终端与边缘计算节点 1 之间分配低时延、大带宽的网络资源，并在边缘计算节点 1 上分配相应的算力资源和存储资源，另一方面，考虑到该 VR 应用的一些行为记录需要上传至个人中心，但此项记录可以是非实时上传，因此算力网络在边缘计算节点 1 和云计算节点之间分配一条 SLA 相对较低的网络资源。而客户 B 用手机终端查看私人视频，需要加密通道，但考虑到手机终端有一定的缓存能力，因此只需要在客户 B 的终端与云计算节点之间建立不保证 SLA 的加密连接即可。

当客户 A 处于移动状态时，比如从靠近边缘计算节点 1 的位置移动到了靠近边缘计算节点 2 的位置上，这时算力网络通过探测与计算发现由边缘计算节点 2 来提供服务更好，此时通过广域网建立一条从客户 A 到边缘计算节点 2 的通道，相应的应用也从边缘计算节点 1 迁移到边缘计算节点 2，从而继续为客户 A 的 VR 应用提供低时延和大带宽的服务。客户位置发生变化后，重新部署算力资源如图 4 所示。

综上所述，通过采用云计算技术与网络领域技术最新的成果，如 SDN/NFV 等技术，算力网络能够为客户提供云、网、边深度融合的整体解决方案，并能够在网络范围内实现灵活可控的算力及各类资源的调度，既能满足客户的高性能要求，又能有效降低建设与维护成本，提升整网运营效率。

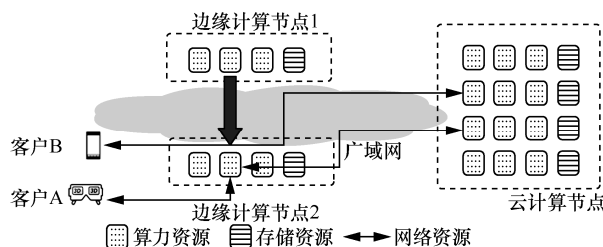


图4 客户位置发生变化后，重新部署算力资源示例

4 面向 AI 应用需求的算力网络实践

算力网络中的算力调度实现可以有多种方案，比如通过集中管控平台实现统一调度，也可以通过分布式路由协议来实现，如在 BBF 计划立项的城域算力网络项目。因此为了进一步验证算力网络的能力，在前期研究的基础上组建了基于 SDN、NFV、AI、云计算等新型技术的试验环境，并且结合 AI 赋能平台，为 AI 应用需求提供灵活的算力调度系统，这套系统被取名为“AI 算力网络”，其架构如图 5 所示。

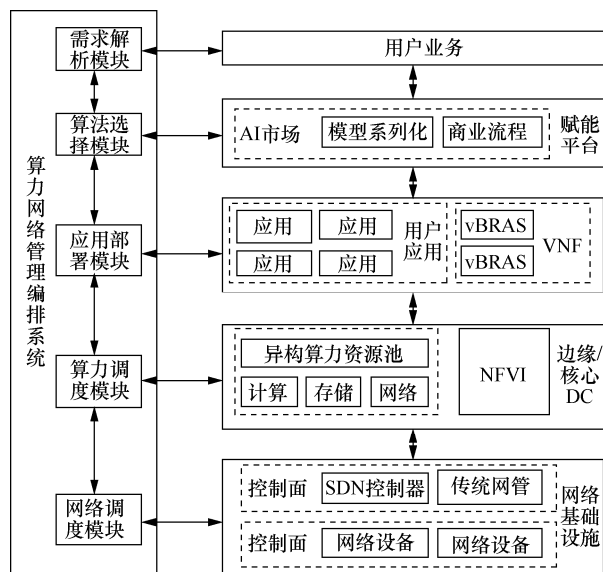


图5 AI 算力网络框架

系统架构中包含的主要部分如下。

(1) 算力网络管理编排系统

算力网络的资源管理和调度系统，根据业务需求对算力资源进行弹性调度，在满足业务实时需求的同时，提高算力利用率。

(2) 赋能平台

为用户的 AI 业务提供基于平台的 AI 服务。

(3) 边缘/核心 DC

包含算力资源基础设施和 NFV 基础设施, 其中, 用户应用部署在异构算力资源池之上, vBRAS、vCPE 等虚拟网关部署在 NFVI 之上。

(4) 网络基础设施

连接用户、边缘云、核心云的网络基础设施, 包括控制面的 SDN 控制器、传统网管以及转发面的网络设备。

算力网络管理编排系统的主要模块功能如下。

(1) 需求解析模块

分析用户业务需求, 根据不同场景将用户业务需求转化为算力资源需求, 根据算力需求划分业务等级, 以确定业务的部署位置、资源等信息。

(2) 算法选择模块

由用户指定或根据需求解析模块的结果, 在 AI 赋能平台中选择用户业务的 AI 服务, 确定用户业务部署的规格。

(3) 应用部署模块

将 AI 赋能平台中的服务部署到指定的节点上。

(4) 算力调度模块

管理核心云和边缘云的算力资源, 根据业务需求为用户分配相应的计算、存储、网络资源, 并根据策略对业务部署位置、业务算力进行弹性调整。

(5) 网络调度模块

管理用户、边缘云、核心云的网络, 在确定用户业务部署位置后, 联合算力调度模块将业务网关下沉到用户业务同一位置, 将业务流量路由到处理节点。

客户可以分级提出不同的需求, AI 算力网络能够自动分析需求, 并分配合适的基础资源。针对 AI 类应用, 客户需求可以细分为以下 4 个指标。

(1) 业务需求

用户 AI 业务所要达到的实际效果, 如处理时延、数据规模等。

(2) 算法需求

针对同种 AI 业务具有多种处理算法, 侧重点不同, 用户可以指定使用何种 AI 算法和模型。

(3) 算力需求

部署用户业务所需要的算力资源, 用户可以指定需要的算力资源规模。

(4) 网络需求

用户业务接入处理节点的网络需求, 用户可以指定接入的网络节点。

针对用户提出的不同场景, AI 算力网络都有对应的处理流程, 具体可以分类以下 4 类场景。

(1) 第一类场景, 客户给出了 4 类需求的指标, AI 算力网络将根据客户输入, 直接分配相应的基础资源并建立连接, 如图 6 所示。

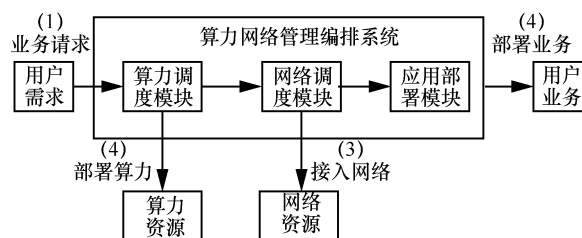


图6 场景一的处理流程

(2) 第二类场景, 客户根据经验选择算法并明确算力资源需求, 但不了解网络资源需求, 因此 AI 算力网络将根据用户需求自动解析用户业务的网络接入点, 并进行业务部署, 如图 7 所示。

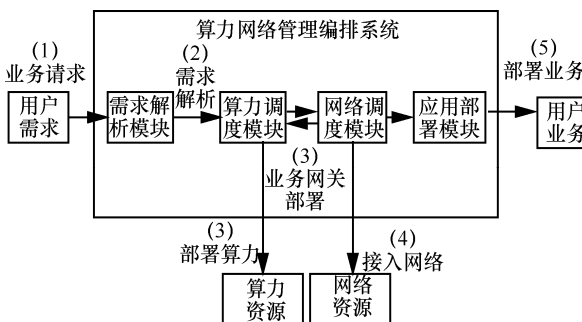
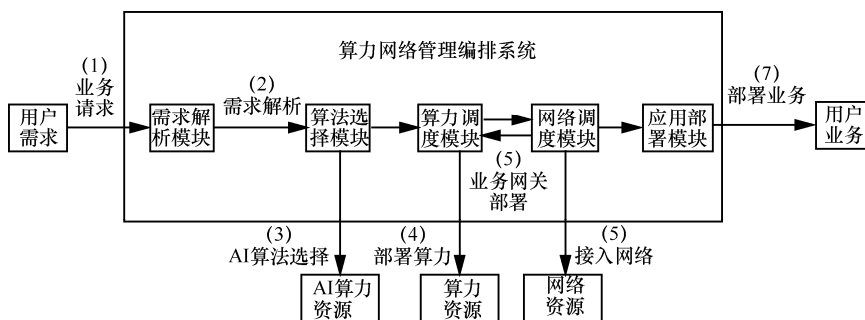
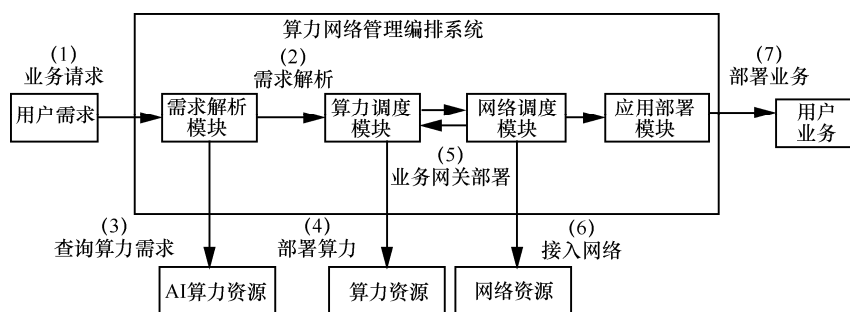


图7 场景二的处理流程

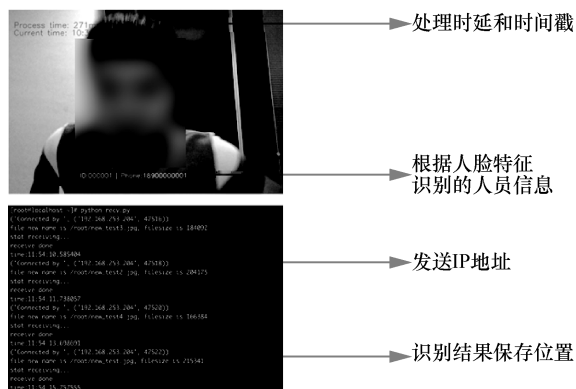
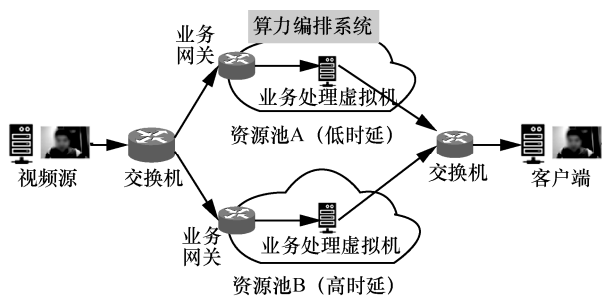
(3) 第三类场景, 客户选择了算法, 但不清楚所需的算力和网络的资源需求, AI 算力网络将



根据 AI 赋能平台中记录的算法所需的算力资源，为用户进行算力分配和网络调度，如图 8 所示。

(4) 第四类场景, 客户只是提出了 AI 应用需求, 因此 AI 算力网络将自动选择最匹配的算法, 并分配对应的算力资源和网络资源, 如图 9 所示。

作为实验验证,在实验室模拟了针对人脸识别业务的算力调度。实验室搭建了两套环境不同的云平台,可以分别用于处理人脸识别业务,识别员工的身份、电话等信息,以图片的形式输出并标记处理时延,用户业务要求处理时延在 1 s 以内,实验拓扑和识别结果如图 10 和图 11 所示。



当用户业务处于闲时, 用户业务将部署到资源池 B 进行处理, 在满足业务需求的同时, 节省高性能资源池 A 的资源; 而当用户处于忙时, 资源池 B 的处理时延将到达 1 s 以上, 难以满足用户需求, 此时需要将用户业务调度到资源池 A 进行处理, 利用 NFVO 在资源池 A 进行业务网关和处理虚拟机的部署。

由于资源池 A 可以提供更低的时延和更多的算力, 满足忙时的业务需求, 将时延重新降低到 1 s

以内,调度前和调度后的结果对比如图 12 所示。



图 12 调度前和调度后的结果对比

5 结束语

边缘计算作为 5G 时代最重要的创新场景,能够为客户提供低时延、大带宽等多种业务保障,但随着研究和部署的深入,边缘计算与云计算、网络(尤其是广域网)之间的协同成为新的研究点。本文针对在多级计算节点之间按需部署与灵活调度算力的需求,提出了一种基于云、网、边深度融合的方案,通过构建算力网络,满足不同类型业务的需求。

参考文献:

- [1] ETSI ISG MEC. Multi-access edge computing (MEC) framework and reference architecture[R]. 2016.
- [2] 边缘计算产业联盟,工业互联网产业联盟(AII). 边缘计算参考架构 3.0[R]. 2016.
ECC, AII. Multi-access edge computing(MEC) reference architecture 3.0[R]. 2016.
- [3] SDN/NFV 产业联盟. MEC 行业应用白皮书[R]. 2018.
Alliance of SDN/NFV Industry. White paper on industry applications of MEC[R]. 2018.
- [4] 阿里云技术有限公司,中国电子技术标准化研究院. 边缘云计算技术及标准化白皮书[R]. 2018.
Alibaba Cloud Computing Co., Ltd., CESI. White paper on technologies and standardization of edge cloud computing[R]. 2018
- [5] 网络 5.0 产业和技术创新联盟. 网络 5.0 技术白皮书(2019)[R]. 2019.
N5A. White paper on network 5.0 technologies(2019)[R]. 2019.
- [6] 百度,中国联合网络通信集团有限公司,中国电信集团有限公司,等. AI 边缘计算技术白皮书(2018-2019)[R]. 2019.
Baidu, CUCC, CTCC, et al. AI-oriented edge computing technical white paper[R]. 2019.
- [7] 云计算开源产业联盟. 云网融合发展白皮书(2019 年)[R]. 2019.
OSCAR. White paper on development of integration of cloud and network(2019)[R]. 2019.
- [8] GAO Y, GUAN H B H, QI Z Q, et al. Service level agreement

based energy-efficient resource management in cloud data centers[J]. Computers and Electrical Engineering, 2013: 1621-1633.

- [9] 李林哲,周佩雷,程鹏,等. 边缘计算的架构、挑战与应用[J]. 大数据, 2019(2): 3-16.
LI L Z, ZHOU P L, CHENG P, et al. Architecture, challenges and applications of edge computing[J]. Big Data Research, 2019(2): 3-16.
- [10] 施巍松,张星洲,王一帆,等. 边缘计算:现状与展望[J]. 计算机研究与发展, 2019, 56(1): 69-89.
SHI W S, ZHANG X Z, WANG Y F, et al. Edge computing: state-of-the-art and future directions[J]. Journal of Computer Research and Development, 2019, 56(1): 69-89.
- [11] 张开元,桂小林,任德旺,等. 移动边缘网络中计算迁移与内容缓存研究综述[J]. 软件学报, 2019, 30(8): 2491-2516.
ZHANG K Y, GUI X L, REN D W, et al. Survey of computation offloading and edge caching in mobile edge Networks[J]. Journal of Software, 2019, 30(8): 2491-2516.

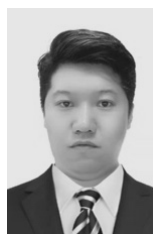
[作者简介]



雷波(1980—),男,中国电信股份有限公司北京研究院新兴信息技术研究所 IP 与未来网络研究中心主任、高级工程师,CCSA“网络 5.0 技术标准推进委员会”管理与运营组组长,主要研究方向为未来网络架构、新型 IP 网络技术等。



刘增义(1992—),男,中国电信股份有限公司北京研究院新兴信息技术研究所 IP 与未来网络研究中心工程师,主要研究方向为网络功能虚拟化、未来网络等。



王旭亮(1986—),男,中国电信股份有限公司北京研究院新兴信息技术研究所 IP 与未来网络研究中心工程师,主要研究方向为网络功能虚拟化、未来网络等。

杨明川(1973—),男,中国电信股份有限公司北京研究院副院长,主要研究方向为云计算、大数据和人工智能、区块链的研究与融合技术等。

陈运清(1964—),男,中国电信股份有限公司北京研究院院长、教授级高级工程师,CCSA“网络 5.0 技术标准推进委员会”副主席,主要研究方向为未来网络架构、新型 IP 网络技术等。