

近端云计算:后云计算时代的机遇与挑战

周悦芝 张迪

(清华大学 计算机科学与技术系, 北京 100084)

(清华信息科学与技术国家实验室(筹), 北京 100084)

摘要 随着普适化智能终端以及泛在化网络技术的快速发展,新型网络应用和服务不断涌现,诸如物联网、智慧城市、智能电网、虚拟/增强现实、无人驾驶等.以集中式计算和存储为主要特征的云计算已经难以满足这些技术和应用的发展需求.近年来,为了克服云计算的局限性,雾计算、移动边缘计算以及边缘计算等近端云计算模式相继被提出并日益受到关注.这些计算模式虽然名称不一,但是其基本思想都是对云计算进行扩展,即将云计算基础设施从距离用户较远的数据中心移至距离用户终端较近的边缘路由器、移动基站或者服务器上,以克服云计算的技术和应用瓶颈.本文对近年来出现的近端云计算模式进行了系统综述,总结了当前几种近端云计算模式的基本特征.首先,分析了云计算在发展中所遇到的主要技术和应用瓶颈.其次,介绍了近端云计算模式的概念和架构,并总结了几种近端云计算模式.接着,讨论了近端云计算所面临的主要发展机遇和技术挑战.最后,对近端云计算的未来发展进行了展望.

关键词 云计算;雾计算;边缘计算;移动边缘计算;多通路边缘计算

中图法分类号 TP301

论文引用格式:

周悦芝,张迪.近端云计算:后云计算时代的机遇与挑战, Vol.41, 在线出版号No.25

ZHOU Yue-Zhi,ZHANG Di,Near-End Cloud Computing: Opportunities and Challenges in the Post-Cloud Computing Era, 2018,Vol.41,Online Publishing No.25

Near-End Cloud Computing: Opportunities and Challenges in the Post-Cloud Computing Era

ZHOU Yue-Zhi ZHANG Di

(Department of Computer Science and Technology, Tsinghua University, Beijing 100084)

(TNList, Tsinghua University, Beijing 100084)

Abstract Cloud computing is a form of network computing paradigm that can provide on-demand access to shared pools of configurable computing resources, such as computation, storage, and services. Since its introduction in 2007, it has attracted a lot of attention and has developed rapidly over the last decade. In recent years, with the rapid development of pervasive smart devices and ubiquitous network technologies, various kinds of new network applications and services are continuously emerging, such as the Internet of things, smart city, smart grid, virtual/augmented reality, unmanned vehicles. However, it is very difficult for cloud computing to meet the diverse requirements of these new network applications and services due to the characteristic of centralized computation and storage. The capabilities of heterogeneous devices and edge equipment are not fully leveraged in cloud computing. Cloud computing also encounters the bottleneck of heterogeneous and long-distance networks. Moreover, the demands of new complex network application scenarios, such as the

Internet of Things and 5G, cannot be satisfied by cloud computing. To overcome the limitations of cloud computing, a variety of new network computing paradigms, such as fog computing, multi-access edge computing, and edge computing are proposed by industry and academia. Although they employ different terms, the basic idea of them is to extend cloud computing and move the computing infrastructure from remote data centers to edge routers, base stations and servers that are close to end-devices, thus overcoming the technology and application bottlenecks of cloud computing. In this paper, we collectively refer to these new network computing paradigms as near-end cloud computing paradigm. To make it easier to understand the underlying causes of the near-end cloud computing paradigm and to grasp the application scenarios and challenges of it, we systematically summarize the emerged near-end cloud computing paradigms in recent years and sum up the basic characteristics of them. First, the main technology and application bottlenecks encountered in the development of cloud computing are summarized and analyzed in detail. Second, the concept and architecture of the near-end cloud computing paradigm are introduced, and several near-end cloud computing paradigms, including fog computing, multi-access edge computing, and edge computing are summarized. This shows that these near-end cloud computing paradigms are not completely different from cloud computing, but rather a natural extension of cloud computing from large-scale centralization to small-scale centralization and distribution, which can be regarded as a historical regression to the distributed computing paradigm. Third, the development opportunities associated with near-end cloud computing are discussed with application scenarios, including the Internet of Things, 5G, smart grid, virtual/augmented reality, and unmanned aerial vehicle. Fourth, the technological challenges including architecture and model, network access and management, resource management and scheduling, pricing and incentive mechanism, security and privacy protection are discussed in details. Finally, the future development of near-end cloud computing is prospected. It is believed that as problems faced by cloud computing become increasingly prominent, near-end cloud computing will undoubtedly become a hot research topic in both the industrial and academic communities in the near future.

Key words cloud computing; fog computing; edge computing; mobile edge computing; multi-access edge computing

1 引言

云计算自从在 2007 年由 IBM 公司提出以来¹, 得到了 IBM、谷歌、亚马逊以及微软等大型商业公司的大力推进. 近几年来, 云计算及其服务在诸多领域得到了广泛应用, 受到了人们的普遍认可, 也产生了巨大的商业价值. 2016 年, 亚马逊云计算业务营收达到了 122 亿美元². 微软预测其云计算业务营收在 2018 年前将达到 200 亿美元³. 近年来, 计算机软硬件和网络通信技术的蓬勃发展, 一方面促进了云计算

的发展, 但另外一方面也使得云计算本身固有的问题和缺陷开始显现出来. 因此, 为了应对云计算发展和应用所面临的问题, 人们已经开始探索后云计算时代的新型网络计算模式.

首先, 普适化智能终端技术发展迅速, 出现了各式各样的新型智能终端设备, 并得到了广泛应用. 以智能移动终端(智能手机等)为例, 截至到 2011 年, 全球智能移动终端的出货量已经超过 PC 终端. 2016 年, 全球移动用户的总量已经达到 70 亿⁴, 我国移动互联网用户在 2016 年突破了 9 亿⁵. 根据摩尔定律, 这些智能移动终端的能力发展迅速, 比如华为 P10 手机, 其 CPU 具有 8 个核: 4 个 2.4GHz 核心和 4 个 1.8GHz 核心, 已经超过一些 PC 终端. 再以智能穿戴设备为例, 智能手表、智能手环等智能穿戴设备近年

¹ IBM Introduces Ready-to-Use Cloud Computing, <https://www-03.ibm.com/press/us/en/pressrelease/22613.wss> 2007,11,15

² Amazon's massive cloud business hit over \$12 billion in revenue and \$3 billion in profit in 2016, <http://www.businessinsider.com/amazons-cloud-business-hits-over-12-billion-in-revenue-2017-2> 2017,2,2

³ Microsoft vows to have \$20 billion in cloud revenue by 2018, <http://www.businessinsider.sg/microsoft-20-billion-cloud-revenue-and-1-billion-windows-10-pcs-by-2018-2015-4/?r=US&IR=T#iz25lv5lyAbEVkd3.9> 2015,4,29

⁴ ICT Facts and Figures 2016, <http://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2016.pdf> 2016,6,1

⁵ Economic operation of Telecom industry November 2016, <http://www.miit.gov.cn/n1146285/n1146352/n3054355/n3057511/n3057518/c5426194/content.html> 2016,6,1

来也得到了飞速发展.此外,城市、社区、甚至山河湖泊也都普遍部署了各种具有不同处理能力的智能或非智能传感设备,如雷达、摄像头、水质/火灾传感器等.这些普适化智能或非智能的终端,在处理和存储、网络连接以及能耗等方面差异非常大.集中式的云计算是否能够适应这些移动终端的特点并最大潜力地发挥这些不同终端设备的能力,这是云计算普遍应用需要考虑的难题.

其次,泛在化网络技术得到了进一步快速发展和部署.4G-LTE、LTE-A、Wi-Fi等无线宽带技术已经得到广泛使用,5G和D2D(Device to Device)技术正在迅速发展,有望在未来几年得到商业部署.无线宽带技术的发展,虽然加快了用户访问云服务的速度,提高了用户体验.但是,用户的访问速度和需求还仍然得不到满足.以在线视频应用为例,移动在线视频播放不但要消耗大量带宽,而且网络延迟对用户体验的影响极大.旨在提高用户带宽和访问速度的5G和D2D技术的发展,既为云计算带来了新的发展动力,同时也给云计算能否充分利用不断更新的网络设备和技术,进一步提高用户体验带来了新的挑战.

与此同时,随着软件定义网络SDN(Software Defined Networking)和网络功能虚拟化NFV(Network Functions Virtualization)等新兴网络技术的发展和应用,边缘的网络接入设备甚至核心路由设备的处理和存储能力不断地增强,而且其功能可以根据用户需求进行扩展和增值.这与集中式的云计算理念不完全一致,也为云计算的进一步发展和应用带来了挑战.

进一步,普适化智能终端和泛在化网络技术的快速发展,也催生了丰富多彩的新型网络应用和服务.物联网(Internet of Things, IoT)、车联网(Internet of Vehicles, IoV)、万物网(Internet of Everything, IoE)、智慧地球(Smart Planet)、智慧城市(Smart City)、智能电网(Smart Grid),以及社交网络(Social Networks)、未来网络(Content/Media/Service Oriented Networking)等这些应用也逐渐被大众所熟知.近年来,虚拟现实(Virtual Reality, VR)/增强现实(Augmented Reality, AR),无人驾驶(Self-Driving Vehicle/Unmanned Aerial Vehicle)等新型网络应用和服务不断涌现.集中式的云计算模式能否满足所有这些新型网络应用和服务快速增长的不同需求,这是云计算发展面临的另外一个重要挑战.

综上所述,不难看出,普适化智能终端、泛在化

网络技术以及新型网络应用和服务日新月异的发展,使得以集中式计算和存储为根本性特征的云计算模式难以普遍适应诸多的技术和应用场景的需要.为此,产业界和学术界从2011年就开始了对于后云计算时代网络计算模式的思考和探索^[1],并且取得了初步的成果.雾计算(Fog/Mist Computing)^[2]、移动边缘计算(Mobile Edge Computing, MEC)^[3]、边缘计算(Edge Computing)^[4]等多种新型网络计算模式被相继提出来并开始得到初步研究和应用.鉴于网络计算模式发展对整个计算机甚至通讯产业的发展具有极其重要的作用和意义,本文试图对后云计算时代的网络计算模式进行初步的分析和总结,以推动这一领域的研究和发展.

当前的几种新型网络计算模式,虽然被提出时所考虑的技术和应用的出发点有些差异,但是其基本思想和核心理念是一致的:即都是试图将云计算中心的设备部署在物理或逻辑上距离终端和用户较近的基础设施上,从而利用这些较近的基础设施所拥有的计算和存储等资源来完成终端和用户想要完成的计算、存储和网络等各类任务.因为这些附近的基础设施一般具有更低的延迟,所以任务的响应时间可以很大程度上被降低,相应的也就可以提高用户的体验.因此,为了表述方便,同时也为了概括计算和存储等任务在距离终端和用户较近的基础设施上完成这一特点,本文将最近提出的新型网络计算模式统一地称为“近端云计算”(Near-End Cloud Computing).相应地,因为距离终端和用户较远,传统意义上的云计算被称为“远端云计算”.本文将首先从更加广泛的意义上对近端云计算的体系结构进行总结分析,以便于理解和洞察近端云计算的基本内涵,从而把握近端云计算与传统云计算的区别.另外,为了便于读者深入了解各种近端云计算模式的思想概念以及它们之间的区别,本文将详细地介绍几种近端云计算模式,包括雾计算、移动边缘计算、边缘计算等.这些近端云计算模式分别由网络设备厂商Cisco公司、欧洲电信标准化协会(European Telecommunications Standards Institute, ETSI)和学术界提出,在基本理念一致的前提下,它们各有侧重点.因此,本文也将对这些近端云计算模式的侧重点进行分析和比较.

无论是产业界还是学术界提出的近端云计算模式,它们能否得到广泛的接受和发展,取决于诸多因素的综合作用.虽然目前很难预测哪一种近端云计算模式将占据主导地位,但是可以肯定的一点是:

近端云计算在不远的将来将会得到大力的发展.然而,近端云计算的发展在面临许多机遇的同时,也面临着诸多技术挑战.因此,本文还试图从不同角度对近端云计算的发展机遇和技术挑战作一个初步分析和总结.

本文第2节对云计算发展遇到的问题进行了概要分析,揭示了云计算的基本特征及其发展演化的必然趋势.第3节介绍了最近出现的近端云计算模式,并总结了这些近端云计算模式的基本特征.第4节较详细地分析、比较和总结了几种近端云计算模式.第5节通过一些具体的应用例子,总结了近端云计算发展的机遇.第6节讨论了近端云计算发展将面临的一些技术挑战.最后展望了近端云计算的未来发展.

2 云计算的发展与问题

随着计算机软硬件技术和网络通信技术的蓬勃迅速发展,计算模式也经历了不同时期的变革.回顾历史,最初的计算模式为大型主机模式.在大型主机模式中,计算和存储等任务通过主机直接完成,而终端设备仅负责输入和输出的交互任务.因此,从本质上来说,大型主机模式是一种集中式的计算模式.随着计算机小型化和普及化的不断发展,计算模式从大型主机模式跨入了PC计算模式.不同于大型主机模式,PC计算模式中的计算和存储等任务可以分散在不同的PC计算机上来完成,而这些PC计算机又是相互独立的.所以,PC计算模式实质上是一种分散式的计算模式.随着网络技术在上世纪60到70年代兴起和发展之后,计算模式又再一次踏入了网络计算模式的时代.通过网络技术,原本相互独立的PC计算机能够相互连接起来,并且还可以与具备特殊功能的服务器一起完成指定的任务.在网络计算模式时代,有很多网络计算模式被相继提出,包括网格计算^[5]、服务计算^[6]、自主计算^[7,8]、透明计算^[9]和云计算等.目前,云计算是吸引了普遍的关注并且得到了大范围的应用.云计算的基本特征是将原本分散在很多PC计算机上的计算和存储等任务通过集中式的数据中心来处理.尽管从表现形态和技术特征来看,云计算与大型主机之间差异很大,但本质上云计算延续了大型主机的集中式处理特点,是集中式计算模式的螺旋式的历史轮回.

追溯计算模式的发展历程,可以看出无论是集中还是分散的模式,它们都有本身独特的优劣势.在

不同的历史时期,由于应用场景需求的变化和技术发展的不同特点,它们分别取得不同程度的主导地位,并且这种发展呈现出了一种螺旋式循环的历史规律.目前,虽然云计算取得了比较大的成功,但随着应用场景需求和技术发展趋势的变化,云计算本质上的集中式特征使得其本身固有的缺陷和瓶颈逐渐表现出来.下面将从三个方面对云计算发展过程中面临的问题进行分析总结.

2.1 异构终端和边缘设备能力的利用

首先,随着普适化智能终端的发展,出现了各式各样异构的智能或非智能化的网络终端设备,如智能手机、智能可穿戴设备(手表、手环、眼镜等)、智能电器、智能仪表、智能传感器等大小、形状、能力不一的各类终端设备.一方面,由于摩尔定律的持续作用,虽然这些终端设备的体积变得越来越小,但他们的计算和存储等方面的资源却变得越来越强大.这就在一定程度上导致了终端设备资源的过剩和闲置.另外一个方面,其他一些终端设备的计算和存储资源由于应用场景或者技术的要求等原因则显得仍然无法满足需求,例如物联网智能传感器的计算和存储等方面的资源总表现的明显不足.在云计算中,由于终端设备只是进行输入和输出的人机交互任务.因此,一些终端设备较强的计算和存储资源往往不能得到充分地利用,这也就在一定程度上造成了终端设备资源的浪费.更重要的是,在云计算中,其他一些自身资源不足的终端设备却无法通过利用附近其他具有空闲资源的终端设备来扩展自身有限的资源,从而使得它们能够完成需要消耗较多资源的任务.因此,从多个终端设备的整体来说,终端设备的计算和存储等资源无法得到有效的利用.

其次,目前的智能终端,尤其是智能移动终端,大都装备或具有某种或多种传感器,能够感知自身的位置(如GPS)或者周围的环境状态.这些位置或者环境信息的采集,将产生大量的传感数据.如果将这些超大规模的传感数据均上传到云计算的大型数据中心进行处理,不但费时费力,甚至会因为数据处理不及时而变得失去利用的意义和价值.

再次,随着硬件技术的发展,边缘网络设备能力也变得更强大.边缘网络接入或者路由设备的处理和存储能力的增强,使得用户利用边缘网络设备来实现自己所需要的增值服务已经完全成为一种可能.通过软件定义网络和网络功能虚拟化技术,网络边缘设备的能力可以得到充分发挥.在5G网络中,

基站的增值服务功能得到进一步扩展,用户可以利用基站提供的专门服务器来定制和实现自己所需要的计算、存储或者通信等功能。

2.2 异构与长距网络的瓶颈

随着泛在化网络技术(Ubiquitous Network Technology)的快速发展,不仅有线宽带网络的速度和带宽得到了很大的提升,无线网络更是变得随时随地都可以接入。虽然网络技术的迅速发展一方面为云计算的发展和应用提供了技术支撑和保障。但是,另外一方面,云计算的集中式的处理特性也因为网络技术而限制了其应用的性能和范围。

首先,移动互联网和物联网的快速发展,使得智能移动终端和无线宽带访问已经变得非常普及。移动终端、用户和服务已经成为当前计算和应用服务的主流,这使得云计算在移动用户访问场景下受到挑战。例如,移动终端用户希望在不同地域、不同时间和不同服务提供者之间能够得到无缝的服务访问体验,这对于缺乏服务提供商之间相互操作支持的云服务而言是难以有效支持的。

其次,在传统的云计算模式中,终端设备和集中的云计算数据中心之间通常需要跨越远距离的广域网络以及各类异质的无线网络。虽然长距离广域网络的带宽和速度在一直被改善^[10],但由于网络内容的快速增长所造成的网络拥塞以及网络数据包处理而造成的延迟等原因使得长距网络的性能和可靠性保障仍然不能满足很多应用的需求。更严重的是,因为网络技术本身所需要的路由路径选择、网络路由节点的数据包处理以及网络内容安全监管等原因使得长距网络的的性能问题很难在短期内得到解决,而且甚至在未来很长的一段时期也可能无法得到明显的改观。另外,无线网络由于其技术本身固有的广播特性,使得其很容易受到干扰或者被网络容量的影响,也就导致了无线网络的服务质量也很难得到有效地保障。

据文献[11]可知,在交互式应用中,如果响应延迟达到 150 毫秒以上,人们就会感觉到不满意,在 1 秒以上人们就会感到非常的厌烦。在云计算中,计算和存储等任务都是在远距离的数据中心中进行集中处理,而长距网络的性能和延迟问题以及无线接入网络的干扰等问题使得云计算很难满足交互式实时应用的需求。目前云计算应用的现状似乎也验证了这一点问题。当前,从终端用户来看,日常云计算应用较多的是云存储,例如百度云盘、360 云盘、腾讯微云等,以数据/文件存储和共享为主要应用。而云

桌面等云计算典型应用则部署和应用的非常少。因此,长距离的网络瓶颈,使得云计算的应用范围和效果被限制。

2.3 新型网络应用和服务的需求

随着泛在化网络技术和普适化终端的快速发展,各类新型的网络应用和服务层出不穷,令人应接不暇。例如,物联网、5G 网络、智能电网、无人机等。这些复杂化的新型网络应用和服务在终端设备的能力、延迟和性能需求等要求各有差异,这使得云计算的在广泛的适用性方面面临很大的技术挑战。下面,本节将以物联网和 5G 网络为代表来总结分析云计算在适应新型网络应用和服务方面所存在的不足之处。

首先,以物联网应用场景为例进行分析。物联网应用场景实际上涵盖了一系列应用和服务,如车联网或车车网(Connected Vehicle)、智慧电网、智慧城市,以及更加广泛的无线传感器和执行器网络(Wireless Sensors and Actuators Networks, WSNs)等。云计算主要在以下几个方面无法满足物联网应用场景的需求:

- 无线接入:如前所述,无线接入不稳定的性能和延迟特性,使得云计算的集中式处理很难满足那些低延迟应用需求,如游戏(Gaming)、视频流(Video Streaming)、增强现实等。
- 移动支持和位置及环境感知:在物联网应用中,节点可移动,并且具有位置和环境感知功能,这些都要求应用能够具有利用这些位置和环境信息进行及时动态调整的功能。
- 传感节点数据的及时和快速处理:物联网,例如智能电网中,具有大规模传感节点和大规模地理分布的传感器网负责收集来自应用和自然环境的大数据,这些数据需要进行及时和快速的处理,而不是云计算擅长的批数据处理。
- 节点异构、互操作以及协作:一些物联网服务,如视频流或数据流,要求在异构节点间进行无缝连接和迁移,这可能就需要节点之间进行互操作和协作,以提供更好的服务。
- 实时处理:在带有执行器或控制器功能的应用中,来自传感器的数据必须在保护或者控制功能决策中得到实时处理,否则就会造成重要的经济和社会损失。

其次,以 5G 应用场景为例进行分析。5G 网络的

主要特点是:范围更广、带宽更宽(Gigabit)、延迟更短(毫秒级)、软件定义无线电、D2D 通信等.因此,5G 网络可以支持的应用范围更广,除了 IoT 和 IoE,也将可以支持一些任务关键服务(Mission Critical Service)、工厂制造、政府决策、教育、医疗等多种多样的应用.这些需求也给云计算带来了极大的挑战.简要分析如下:

- 内容高效分发:数据分发,尤其是媒体内容分发,占据和使用了互联网的大部分带宽.由于媒体内容的同一性,使得集中式的分发模式实际上造成巨大的带宽浪费,也给支持更大规模的用户、更优质的大规模数据分发服务带来困难.
- 代码 / 计算 / 数据 / 网络 卸载 (Code/Computation/Data/Networking Offloading):由于智能移动终端和可穿戴式设备在计算、存储以及网络能力等方面受限,因此如何利用云计算或者周围其它相邻设备的资源来辅助完成其自身所需要的计算、存储或者联网功能,是一条克服移动终端资源受限的良好途径.然而,由于这些卸载的用户体验严重依赖于延迟,因此这些卸载不太适合部署到远端云计算中心来执行.
- 移动大数据分析:智能移动设备的大量应用产生了大量数据,对这些大数据的分析和处理可以对企业市场和产品应用进行精准的定位.传统的依赖于大型数据中心的批量或者流式数据处理将带来高带宽和延迟.
- 智能交通/政务/医疗:如果这些智能领域应用,如交通信息发布、流量调度、事故和应急处理等都在远程数据中心进行集中处理,则很可能会因为计算或传输延迟而错过最佳处理时机.
- 端端协作与应用:通过 5G 在无线电级别提供的 D2D 功能,端端实时通信和协作成为可能.利用邻近节点和服务 (Proximity node/service)提供的上下文环境信息,道路交通事故和远程机器人急救等应用将可以提供更精准快速的处理.

从上述分析可以看出,集中式的云计算,由于网络传输和延迟瓶颈的存在,很难满足终端异构化、普适化和网络高速化、泛在化条件下涌现的复杂应用和服务的需求,尤其是物联网/万物网以及各类实时交互式应用等.这就在一定程度上延缓了新型网络

应用研发的速度.因此,探索和研究新的网络计算模式以弥补云计算存在的不足已经成为一种亟需.

3 近端云计算的兴起

云计算固有的集中式处理特性,并不能很好地满足日新月异的普适化终端、泛在化网络技术和新型网络化应用和服务的发展需求.因此,产业界和学术界已经开始对计算模式进行新的探索和研究.从 2011 年 Cisco 公司提出雾计算以来,后云计算时代相似计算模式理念以不同的名词形式被相继提出,如移动边缘计算、露计算和边缘计算等.

这些新型网络计算模式虽然是由不同的组织从不同的出发点和角度提出来的,但是它们具有相似的基本原理和思想,也即“拉近云计算中心与终端用户之间的距离”.具体来说,原本集中部署的距离终端和用户较远的数据中心被延伸到了距离较近的网络边缘.这在一定程度上解决了长距网络和异构无线网络延迟的影响¹.从具体的技术实现来说,网络边缘不仅可以边缘路由器、通信基站和服务器等,也可以是专门部署的服务器或者小型的数据中心等,甚至也可以是其他具有有空闲资源的终端设备.很显然,这些新型网络计算模式并非与云计算迥乎不同,而是云计算由集中式处理模式向分散式和小型集中式模式的一种自然延伸.从历史上看,这其实是分散式 PC 计算模式的历史轮回.

为了方便分析不同的计算模式的特点,以及它们之间的演变规律,在近端云计算模式尚未有统一明确定义的前提下,本文试图从体系结构层数、计算和存储方式、执行地点和顺序等方面比较和分析不同的计算模式.表 1 从上述几个方面比较了大型主机、PC 和云计算等计算模式.

表 1 不同计算模式的比较

计算模式	体系结构层数	计算或存储方式	计算或存储执行地点	计算或存储执行顺序
大型主机	一层	集中式	主机	串行
PC 计算	一层	分散式	PC 终端	并行
云计算	二层	集中式	大型数据中心	串行
雾计算 / MEC / 边缘计算	三层或更多层	小型集中式	路由器、通信基站、边缘服务器、小型数据中心等	串行或并行

¹ 其实除了将云计算设施部署在离终端更近地点的外,近年来还出现了其它扩展的方式,如移动云、混和云、联邦云、以及自组织云等.

如图 1 所示,云计算的体系结构包括终端设备和云计算中心两层.而新兴的网络计算模式区别于现有云计算.以雾计算为例来说,根据 OpenFog 联盟对雾计算的定义(参见 4.1 节),可以知道雾计算的资源和服务部署在云到物之间的任何地方.因此,雾计算在云计算中心和终端设备中间扩展了一层甚至多层雾计算中心或者服务器,将云计算的两层架构延伸到了三层甚至更多层^[12].这些中间层次的雾计算中心或服务器可以根据需要采用不同的结构或者规模来实现不同的功能.例如,在智能电网中^[2,12],电网传感设备采集到的数据首先会传输到微电网进行处理.在此之后,电网数据被传输到小规模的电站或者更大的电网进行进一步的处理.最后,传感数据被传输到云计算中心进行处理.电网传感设备和云计算中心之间可以根据地域范围和行政管理的需要部署多级雾计算处理中心.

在云计算中,所有计算和存储操作都在云计算数据中心上进行,终端只是向云发出请求并接收和展示处理结果.从这个意义上来说,云计算中计算和存储的执行方式是一种集中式的方式,与 PC 计算中计算和存储在每台 PC 上分散进行相对应.雾计算集中了附近的终端节点的计算和存储处理,因此是一种小规模集中式计算模式.

顺序上看,其是一种串行执行的顺序.而在雾计算中,由于一层甚至或者多层雾计算中心的存在,再加上远端的云计算中心,使得任务可以在多个雾计算中心上进行并行处理,因而存在一种并行执行的顺序.

雾计算中的多层架构与混合云中的公有云和私有云组成的混和结构相区分.因为在混合云中私有云由企业自己负责运营和使用,而雾计算基础设施由不同的第三方负责运营并主要提供给移动用户使用.

通过上面的分析,考虑到端和云两个极点,为了便于描述,本文将这些最近提出的几种新型网络计算模式统称为“近端云计算”.它从三个方面对云计算模式进行扩展:一是从体系结构上,如图 1 所示,将云计算的 2 层架构提升至多层架构,即在端和云之间可以设置多个不同的近端或者远端云计算中心;二是计算、存储、网络连接等任务除了可以在近端或远端网络/计算设备或者微型计算/数据中心上执行外,还可以在终端本身或者周围邻近终端上进行;三是除了小型集中式之外,多个终端设备还可以通过共享资源并形成对等的小范围的协作网络,从而能够完成单个终端设备无法完成的更加复杂任务.

为了方便不同读者对近端云计算和传统云计算的区别进行更加深入的了解,表 2 对近端云计算和云计算的不同进行了比较^[13,14].

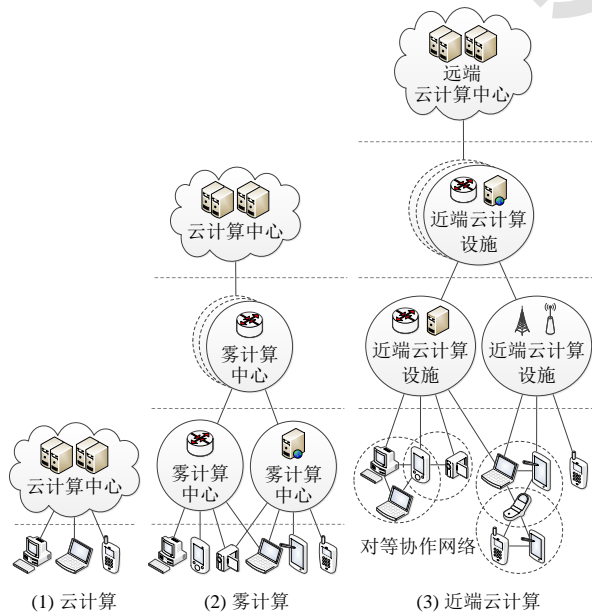


图 1 云计算、雾计算和近端云计算的体系结构

从计算和存储的执行顺序来看,在云计算中,所有的任务均需要在集中的云计算数据中心进行处理.因此,只有从云计算数据中心获得处理结果之后,终端设备才能开始执行接下来的任务.因而,从终端和云计算中心之间涉及的计算和存储任务的执行

表 2 近端云计算与传统云计算的比较

参数	传统云计算	近端云计算
目标用户	普通互联网用户	主要为移动用户
服务类型	从全球范围内收集到的信息	特定部署环境的本地信息服务
硬件	丰富且可扩展的存储空间和计算能力	有限的存储和计算能力
距离用户的距离	在物理上距离用户很远	在物理上很接近用户
工作环境	具有空调系统的仓库大小的建筑	室外或者室内
支持公司	大型互联网服务公司	运营商和网络设备制造商等
延迟	高	低
延迟抖动	高	非常低
网络接入方式	固定和无线	主要是无线
客户端与服务 器之间的距离	多跳	一跳或多跳
位置感知	否	是
移动支持	有限支持	支持
控制方式	集中/分层(完全控制)	分散/分层(部分控制)
服务接入方式	通过中心	在边界或者手持设备

可用性	99.99%	不稳定/高冗余
用户/设备数	数千万到数亿	数百亿
服务节点数	很少	非常多
软件虚拟架构	中心协作的服务器	用户设备

从上述的比较和总结可以看出,近端云计算并非传统云计算模式的简单扩展,其在体系结构模型、服务任务执行方式、任务执行地点、服务质量保障等很多方面都区别于传统的云计算.它是传统云计算模式从集中式计算模式到分散式和小型集中式的自然提升和演化.然而,也正是由于这些根本性区别的存在,可以预见近端云计算模式在面临巨大发展机遇的同时,也将面临着来自商业和技术方面的巨大挑战.本文在对当前几种近端云计算模式做进一步介绍后,也将对近端云计算的机遇和挑战进行初步的分析,以促进该领域相关研究和应用的发展.

4 几种近端云计算模式

从第3节的分析可以看出,目前出现的近端云计算模式在目标用户、目标场景、以及体系结构和实现技术等方面并不相同.为此,将在这一节较详细地介绍和分析几种近端云计算模式的背景、定义以及相关研究进展.

4.1 雾计算 (Fog/Mist Computing)

雾计算的提出主要是为了克服云计算在物联网场景中的不足^[2].如第3节中所述,物联网设备分布的广泛性以及数据处理所需要的实时性使得传统云计算的集中式处理和存储面临巨大的挑战.

雾计算的概念最早由网络设备厂商 Cisco 公司在 2011 年提出^[1],在 2012 年 MCC(Mobile Cloud Computing)会议上以论文形式进行了较详细的阐述^[2].该论文提出的雾计算定义如下:

定义 1. 雾计算是一个高度虚拟化的平台,在终端设备和传统云计算数据中心之间提供计算、存储和网络服务,通常但又不完全位于网络边缘.

在该文中,作者还给出了 IoT 应用的理想模式和计算体系结构(三层结构),如图2所示.最底层是由百万量级的智能物体(传感器等嵌入式设备或系统)组成的智能物联网层,主要完成数据和信息的采集.中间是由数万智能雾服务器组成的边界服务层,完成数据的初步处理.最上层是由数千服务器构成的核心服务层,最终完成数据的核心分析和应用.从这

里可以看出,雾计算是云计算的一种扩展.当然,雾计算也并不是云计算的简单扩展和应用.为了适应物联网应用所需要的位置服务、上下文感知等特征,雾计算引入了一些特殊的性质和特征,比如:1)位于网络边缘位置,位置感知,低延迟;2)地理位置的广泛分布;3)支持内在需要分布式计算和存储资源的大规模传感器网和智能电网等;4)支持节点异构和移动;5)支持超大规模的网络节点;6)实时交互;7)无线访问为主;8)支持服务提供者的互操作和联盟;10)支持在线分析以及与后端云计算的结合.



图2 物联网和雾计算

从上述分析可以看出,Cisco 公司初始给出的雾计算定义是比较粗糙的,它只是对云计算在物联网应用中遇到的问题进行了分析,并简单地将云计算的虚拟化平台移动到网络边缘设备上,进行扩展性的应用.

为了推动雾计算的研究和产业发展,2015年11月,ARM、Cisco、Dell、Intel、Microsoft 以及 Princeton University Edge Computing Laboratory 联合发起成立了 OpenFog 联盟(OpenFog Consortium)¹,现已有来自 15 国家的 53 个成员.OpenFog 联盟对雾计算的定义如下:

定义 2. 雾计算是一个系统级的水平体系结构,它将计算、存储、控制和网络的资源和服务部署在云到物之间的任何地方.

其主要特点如下:

- 水平体系结构:支持多个行业垂直和应用领域为用户和公司提供智能和服务.
- 云到物体的连续服务:将服务和应用程序部署到更接近物体,以及云和物体之间的任何

¹ OpenFog, <http://www.openfogconsortium.org>, 2017,7,10

地方.

- 系统级:从物体,到网络边缘,再到云,跨多个协议层,不只是无线接入系统,也不只是协议层,更不只是端到端系统的一部分,而是跨越物到云之间的系统.

从上述定义可以看出,该定义还是将雾计算的应用主要集中在物联网或万物网,只不过在端与云的连接通路上进行了服务提供的连续化.

需要提醒读者注意的是,目前关于雾计算并没有一个统一的达成共识的定义,因此本文给出的定义,只是雾计算初始提出时给出的定义,其他研究人员和组织还对雾计算的含意进行了自己的阐释或者延伸,由于篇幅的关系,不在本文中一一赘述,有兴趣的读者可进一步参阅文献[14,15,16].

从雾计算的提出和发展,可以看出,随着物联网应用的普遍增多和发展,满足不同物联网应用场景需要的各种智能雾服务器将会被开发出来并得到应用.鉴于物联网设备的分散性,这些智能服务器将具备多种传感器,并将通信、计算和存储等功能集成为一体,可以独立为物联网应用提供场区范围的无线网络接入服务、位置服务、数据缓冲和中继服务等.通过智能雾服务器之间的协作和沟通,为各种物联网应用提供强大的支撑平台.

随着通信、计算和存储技术的发展,智能雾服务器体积和功能的集约化将会得到发展,使得智能雾服务器的部署变得更加广泛,就像现实中的雾一样,弥散在智能物体周围,为其提供便捷快速的各式服务.

4.2 多通路边缘计算 (Multi-access Edge Computing)

多通路边缘计算(Multi-access Edge Computing, MEC)是 2016 年 9 月由移动边缘计算(Mobile Edge Computing, MEC)演变和改名而来¹.移动边缘计算提出的初衷是针对移动蜂窝场景,尤其是 5G 应用场景下如何克服云计算面临的挑战而提出的.随着移动边缘计算概念的发展,访问通路已经不仅仅局限于移动蜂窝网络,还包括 Wi-Fi 网络以及有线网络.因此在保持英文 MEC 缩写的基础上,ETSI 将其概念改名为多通路边缘计算.

MEC 的概念最早由 ETSI 在 2014 年 9 月提出^[3].ETSI ISG MEC (Industry Specification Groups,

Mobile Edge Computing)的倡导者包括: Nokia、Huawei、IBM、Intel、NTT DoCoMo、Vodafone. 现今成员已有 53 家,包括中国的华为和中兴公司.与雾计算的提出不同,MEC 主要是由通信设备厂商提出和倡导的近端云计算模式,因此其目的当然会侧重对通信和应用的支持.MEC 的初始定义是:

定义 3. MEC 在移动网络的边缘为应用程序开发者和内容提供商提供云计算功能和 IT 服务环境.这种环境的特点是具有超低延迟和高带宽,并且应用程序可以实时访问无线网络信息.

其商业上的主要特点有:

- 基于创新和商业价值的新价值链和充满活力的生态系统.
- 运营商可以向第三方授权组织开放其无线接入网络边缘,允许他们灵活快速地部署创新应用和服务.
- 为移动用户、企业和垂直细分市场提供新的创新应用和服务.

很显然,ETSI ISG MEC 给出的 MEC 定义主要侧重于商业和利益方面,对于技术内容没有过多地涉及.MEC 主要是针对移动通信场景需求提出的,旨在联合通信和 IT 等信息产业,在无线接入网络 RAN (Radio Access Network)中增加 IT 和云计算服务能力,为用户提供更多的可供选择的增值服务.

如图 3 所示,MEC 提出时主要是针对 3G/4G/5G 等蜂窝网络应用场景,同时也支持有线网络和 WiFi 等无线网络,其潜在应用包括计算卸载、协作计算、Web 服务内容优化、内容缓存和快速访问等^[17].与雾计算类似,MEC 可以为连接基站的移动终端提供低延迟的访问服务、位置感知服务、就近的计算和存储服务.从分布上说,MEC 服务器主要部署在网络边缘的基站中.这一点与雾服务器可以连续部署在从智能物体到远端云计算中心的路径上不同.

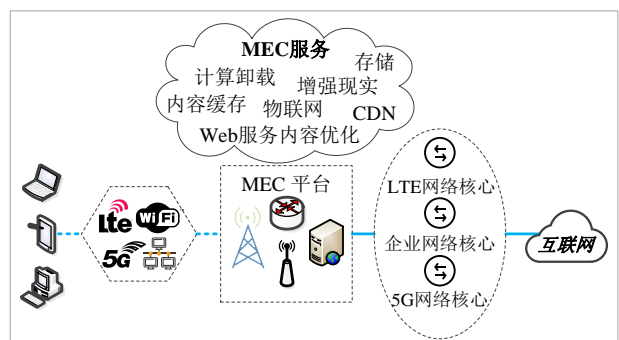


图 3 MEC 架构和应用示意图

为了实现 MEC,可以在 LTE Macro Base Station

¹ ETSI Drops 'Mobile' From MEC, [http://www.lightreading.com/mobile/mec-\(mobile-edge-computing\)/etsi-drops-mobile-from-mec/d/d-id/726273](http://www.lightreading.com/mobile/mec-(mobile-edge-computing)/etsi-drops-mobile-from-mec/d/d-id/726273) 2016,9,21

(eNB)、3G Radio Network Controller (RNC) 站点,或者多种通信技术(3G/LTE)蜂窝聚合站点(Multi-technology [LTE/3G] Cell Aggregation Site)部署 MEC 服务器.由于多技术蜂窝聚合站点既可以安装在企业内部(室内),又可以安装在特定的公众场合(室外)进行特定的场景覆盖,因此可以从基站集群中直接为用户提供本地的快速服务.ETSI ISG EMC 给出的 MEC 服务器平台包括一个 MEC 托管基础设施(MEC Hosting Infrastructure)和一个 MEC 应用平台(MEC Application Platform)^[3].

MEC 托管基础设施包括底层的硬件资源以及虚拟化层,当然也包括到无线网(eNB 或 RNC)或网络的连接.但是该体系结构并不关心如何将服务器集成在无线接入网络内部.实现托管基础设施的硬件组件和细节对于托管在该平台上的应用而言应该是接口化的和抽象化的.

MEC 应用平台,由应用虚拟化管理器(Application Virtualization Manager)和应用平台服务(Application Platform Service)组成,可以为托管的应用提供各种能力.应用虚拟化管理器将基础设施以 IaaS(Infrastructure as a Service)方式为应用提供灵活高效、多租户、运行时和托管环境服务.IaaS 控制器提供安全和资源沙箱功能.应用以虚拟机镜像(Virtual Machine Images)方式打包并传送到 IaaS 上作为虚拟应用(Virtual Appliance)运行.需要注意的是,未来 MEC 也可能支持以 PaaS(Platform as a Service)方式提供服务.MEC 应用平台服务为托管在 MEC 平台上的应用提供一系列中间件应用服务和基础设施服务,具体包括:1)基础设施服务,如通信和服务注册服务;2)无线网络信息服务(Radio Network Information Services)以及流量卸载服务(Traffic Offload Function).

当然,利用 MEC 提供的服务器平台可以支持和提供丰富多彩的移动通信增值服务和应用,比如:主动设备位置、增强现实内容传输、视频分析、RAN 感知的内容优化、分布式内容和 DNS 缓存等.因为篇幅限制,本文不再赘述,读者可以自行参阅 ETSI 网站的内容¹.

随着 5G 技术的发展和应用,5G 基站将集成大量边缘计算服务器,为 5G 网络提供丰富的增值服务,如位置服务、流量卸载、计算卸载、内容传输优化等.由于 MEC 服务器已经具备基站的通信和定位

功能,因此从功能和形态上来看,多通路边缘采用普通服务器的可能性远高于雾计算.由此可以预见到,随着 5G 技术的商业化应用和发展,当前 MEC 服务器所采用的 IaaS 和虚拟机技术,将发展为采用 PaaS 和 SaaS(Software as a Service)技术.与此同时,各种目前人们还未曾想到的移动增值服务将被开发出来,从而为人们提供更加丰富多彩的在线服务.

4.3 边缘计算(Edge Computing)

与雾计算和多通路边缘计算主要由产业界提出和倡导不同,边缘计算^[4]、露计算^[18]、透明计算^[9]等主要由学术界提出和倡导.

边缘计算主要由美国卡内基·梅隆大学(Carnegie Mellon University, CMU)在 2015 年所倡导,旨在从学术研究层面提出一个比雾计算和移动边缘计算更加具有广泛描述意义的计算模式^[4].

2015 年 6 月,卡内基·梅隆大学联合 Vodafone 和 Intel 等公司成立了开放边缘计算倡议组织(Open Edge Computing Initiative).目前该组织包括 CMU, Intel, Nokia, NTT 等 7 家成员.该组织对边缘计算的定义如下²:

定义 4. 边缘计算是一种提供相互连接起来的计算和存储资源的新式网络功能.这种网络功能位于用户所在位置的附近.

边缘计算的主要特点是通过极低延迟提高用户体验,以及通过边缘服务减少数据流量.

文献[19]提出了一种实现边缘计算的体系结构.在该结构中,由移动设备、Cloudlet^[20]以及云三层构成.Cloudlet 位于云和终端设备之间,直接与终端设备通过无线连接的方式连接,广泛部署在互联网架构上,以虚拟机的方式向终端设备提供服务.通过实例化 Cloudlet 上自定义的虚拟机,终端设备以瘦客户端的方式运行和访问所需的应用软件.

另外,从上述定义可以看出,边缘计算与多通路边缘计算类似,实质上只是将移动边缘计算部署的位置从原来的基站扩展到家庭宽带调制解调器、Wi-Fi 无线接入点设备、以及边缘路由器等设备上.

除了边缘计算,学术界还提出了露计算和透明计算,简单介绍如下.

露计算由学术界在 2012 年提出³,在 2015 年以

² Edge Computing, <http://openedgecomputing.org> 2017,7,10

³ The Relationships among Cloud Computing, Fog Computing, and Dew Computing, <http://www.dewcomputing.org/index.php/2015/11/12/the-relationships-among-cloud-computing-fog-computing-and-dew-computing> 2015,11,12

¹ Mobile Edge Computing, <https://portal.etsi.org/tb.aspx?tbid=826&SubTB=826,835> 2017,7,10

论文的形式阐述了露计算的基础架构“云-露”架构^[18],如图 4 所示。“云-露”架构最初提出的目的是为了使得 Web 站点可以在没有互联网连接的情况下也可以被访问.与传统的客户端-服务器架构相比,“云-露”架构增加了露服务器.露服务器是部署在用户本地电脑上的 Web 服务器.在“云-露”架构中,用户的数据不仅保存在云端,还保存在用户的本地.因此,“云-露”架构可以实现在没有互联网连接时的 Web 访问.

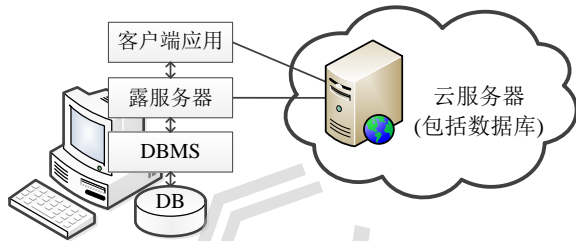


图 4 “云-露”架构示意图

在初始定义的基础上,文献[21,22]又对露计算的定义和层次结构进行了进一步的修订和扩展,将涉及的计算设备从“个人电脑”(Personal Computers)扩展到更加通用的“内部部署的计算机”(On-premise Computers),并将露计算作为云计算和雾计算之下的另一层次.

透明计算是由国内研究人员在 2005 年提出的一种网络计算模式^[9].在透明计算中,距离固定终端较近的透明服务器执行操作系统和应用软件的集

中管理和流式调度运行,因此终端用户可以选择所需要运行的操作系统和应用软件,而不需考虑本地的存储空间和软件安装等管理和维护问题.

文献[23,24]对透明计算的概念和结构进行了扩展,将其从固定终端和固定网络扩展到移动智能终端和异构网络环境,在本地智能终端、相邻终端、近端服务器以及远端云计算服务器之间统一进行资源管理和指令的流式调度执行,克服了云计算单一中心运行模式所带来的问题.

不同于雾计算和多通路边缘计算主要由产业界支持和推动,边缘计算、露计算和透明计算主要由学术界的一些研究人员在推进.因此,相对来说,边缘计算、露计算和透明计算的发展相对缓慢.

综上所述,虽然近端云计算的实质是把云计算基础设施小型化后部署在距离终端和用户较近的地方,但是由于提出时针对的应用场景不同,不同的近端云计算模式在体系结构和实现方式上并不一致.为了使读者清楚地把握这些近端云计算模式的异同,接下来将通过表格的形式详细地比较不同的近端云计算模式.

4.4 几种近端云计算模式的异同

为了便于读者对不同的近端云计算模式的差异有一个总体清晰地把握,表 3 详细地总结了不同近端云计算模式的区别.

表 3 几种近端云计算模式的比较

参数	雾计算	多通路边缘计算	边缘计算	露计算	透明计算
提出者	Cisco ^[1]	ETSI ^[3]	CMU ^[4]	Wang 等 ^[18]	Zhang 等 ^[9]
支持组织	OpenFog 联盟	ETSI ISG MEC	开放边缘计算倡议组织	/	/
启发概念的应用	IoT,无线传感器和执行器网络	Nokia Liquid 应用,电信应用场景,例如内容加速、增强现实等	高速响应、隐私增强以及灵活的云服务,例如增强现实等	互联网 Web 浏览	通过轻量级的设备可访问的异构操作系统和应用
提出目的	满足物联网(IoT)的低延迟、位置感知、移动支持和地理分布的需求	通过将计算和存储从核心网络转移到边缘无线接入网络来减少延迟	在用户的附近提供可连接的计算和存储资源	利用用户的内部部署的计算机以独立云和与云协作的方式向用户提供服务	使用户可以根据需要选择服务,而不用关心服务的安装、维护、管理和安全等
用户	移动用户为主	移动用户为主	移动用户为主	普通互联网用户(包括移动用户)	普通用户和移动用户
接入方式	无线接入为主	无线接入为主	有线和无线接入	有线和无线接入	有线和无线接入
硬件	路由器、交换机、网关等	无线接入点、基站等	路由器、无线接入基站、边缘服务器等	内部部署计算机	用户设备和内部部署的服务器
用户设备利用	是	否	否	是	是
部署环境	物联网和网络边缘	网络边缘	网络边缘	用户设备	用户设备和网络边缘
虚拟平台使用	是	是	是	否	是
服务接入	边界或者用户设备	边界	边界	用户设备	用户设备
典型应用场景	物联网、智能电网	增强现实、智能视频加速	增强现实、人脸识别	互联网 Web 浏览	用户桌面、医疗健康

综合对比几种近端云计算模式可以发现,每种计算模式都有各自的特点.雾计算更加侧重物联网和数据信息的收集分发.移动边缘计算主要从通信运营商的角度出发,侧重于无线接入网络,并没有关注端设备的对等协作.边缘计算则是希望将移动边缘计算进一步扩展到更多网络应用场景.露计算和透明计算主要侧重于终端能力的发挥,以及端设备和云计算之间的协作,而对于边缘路由和无线接入网络并没有给予太多的关注.

从上述对不同近端云计算模式的介绍、分析和比较,不难看出,这些近端云计算模式提出时针对的应用场景、体系结构、特性实现都存在一定的区别.但是随着近端云计算的发展,这些近端云计算模式会在一定程度上相互借鉴,呈现出融合或者互补发展的趋势.由此可以预见:第一,由于不同应用场景具有不同的特征和需求,因此针对不同场景提出的近端云计算模式、体系结构或者实现特性上不会完全统一,这与传统云计算采用大致相同的体系结构和实现特性有所区别;第二,不同的近端云计算模式在发展中会相互借鉴,融合发展,这将使得不同近端云计算模式间具有更多的共同特征,可能会形成通用的基础性近端云计算结构和技术;第三,在具体应用和实现方式上,近端云计算会呈现出更加丰富多彩的样式和风格,并将在通用的基础性架构和技术上衍生出不同的特定实现架构和技术.

5 近端云计算的发展机遇

近端云计算由于其靠近终端设备、分散与集中相结合的计算模式,以及支持终端相互协作等特性使得其在物联网、5G、智能电网、虚拟/增强现实、无人机等诸多应用领域面临巨大的发展机遇.

首先,由于在物理上很接近用户,近端云计算可以克服云计算的响应瓶颈问题,特别是对于一些交互式的应用和用户对于延迟要求比较高的应用来说,近端云计算可以在很大程度上降低网络延迟的影响,从而提高用户体验.其次,近端云计算可以弥补单个终端设备在计算、存储、网络、电池等资源的不足.一方面,近端云计算支持分散与集中相结合的计算模式,因而终端不仅可以利用近端云计算设施的资源,还可以利用远端云计算中心的资源;另一方面,近端云计算支持邻近终端之间的相互协作,这使得终端设备之间可以共享空闲的计算、存储以及传感等资源.因此,近端云计算可以从多个方面克服终

端的资源和能力瓶颈.再次,近端云计算动态灵活的多层次架构,使得其能够适应更多复杂的应用场景.近端云计算通过端与端的对等协作以及近端云计算设施与远端云计算中心相结合的多层次处理架构,可以从延迟、规模、范围等不同的方面满足复杂网络应用的多样化需求,克服云计算在满足新型网络应用场景的不足.

下面通过具体例子来介绍近端云计算的发展机遇.

5.1 物联网

物联网是通过智能和非智能的传感器把各种不同的物体与互联网连接起来,从而实现信息收集、环境监测、健康管理等智能化应用的网络^[25].物联网的基本核心是物体与物体以及人与物体之间的信息交互,它把人们在信息与通信技术世界里沟通的维度从任何时间、任何地点、连接任何人,扩展到了连接任何物体.万物通过互联网连接起来就形成了物联网^[26].

物联网的用途非常广泛,遍及智能交通、环境保护、政府工作、公共安全、环境监测、工业监测、个人监测、敌情侦查、情报收集等多个领域^[26].随着物联网的迅速发展,越来越多的设备接入互联网.据统计¹,截止到2016年底,已经有64亿个“物体”接入互联网.预计到2020年,将有200多亿个“物体”接入互联网.数十亿以前没有接入互联网的设备每天产生着EB级的数据^[27].

如前所述,雾计算就是为解决云计算在物联网领域应用的不足而提出来的.因此,物联网将是近端云计算应用的重要领域之一.以智慧城市为例,通过采集来自各式各样传感器的数据并进行集中分析和利用,物联网可以解决现代城市发展所面临的重大问题,如可持续发展问题.文献^[28]分析了雾计算在可持续化智能城市中的应用前景,包括智慧交通、智慧健康、智慧水和垃圾管理、智慧零售等.利用分布在网络边缘的雾服务器,智慧城市中各种传感器采集的数据可以在网络边缘进行初步分析和及时处理,然后再通过数据聚合,上传到城市数据中心进行进一步的全局处理,为政府机关、社会团体和市民提供信息各种服务和服

其次,近端云计算也可以为车联网提供更好的

¹ Gartner Says 8.4 Billion Connected "Things" Will Be in Use in 2017, Up 31 Percent From 2016, <http://www.gartner.com/newsroom/id/3598917> 2017,2,7

服务.例如,在车联网的路基单元(Road Side Unit, RSU)中植入雾计算处理能力,车与车(V2V)之间产生的交通信息数据在路基单元就可以进行语义层面的数据分析,并提供服务发现和附近联网车辆的管理,而不需要通过远程的云计算数据中心的处理^[29].通过将智能电网与 5G 网络的集成,利用雾计算和云计算基础设施,可以将电动车智能充电等服务请求的时间降低一半多^[30].这种响应时间的缩短,对于便捷的智能充电服务是非常必要的.

此外,物联网中的智能物体,如智能手表等可穿戴设备,其计算、存储和电池续航能力一般都比较弱,不能完成复杂的功能.为了完成复杂的计算和功能,这些智能物体可以将计算任务交给近端云计算服务器来完成,这样既能丰富智能物体的功能,降低服务请求的响应时间,又能降低可穿戴设备的能耗,延长续航时间^[31].此外,利用近端云设施提供的强大的计算和处理能力,物联网网关可以执行先进的认知、学习和其它人工智能算法和功能,例如自动识别物联网应用并进行自适应资源的优化调度^[32].

因此,利用近端云计算可以降低物联网数据处理的规模和能力,提高服务请求的响应速度,增强物联网和智能物体的功能,从而丰富和促进物联网技术的发展和应用.

5.2 5G

如前所述,移动边缘计算是为了克服在移动计算和 5G 应用场景下云计算的不足而产生的.因此 5G 将成为多通路边缘计算等近端云计算模式发展与应用的重要领域.

2017 年 4 月, 3GPP (3rd Generation Partnership Project)在技术规范文档中把支持边缘计算作为 5G 系统的高层特征^[33].这意味着边缘计算将成为 5G 标准和规范的一部分,在未来的移动蜂窝基站(eNB 或 RNC)中将内置边缘云计算服务器或微型数据中心.结合 5G 网络本身的高带宽和低延迟特性,边缘云计算可以为端到端应用提供实时处理和服务,在端和云之间最大化事务处理的速率.同时,本地通信和处理可以提供稳定的性能以及较高的隐私和安全性.这些都极大提高 5G 网络的增值服务能力,衍生出更多丰富多彩的应用.

以在线直播场景为例,在现场可以部署拍摄角度不同的摄像机.很显然,这将产生大量的视频数据,要在目前的 4G 网络中将这些数据通过远程云数据中心处理之后再传输给现场或者非现场观众是非常困难的^[34].在 5G 网络中,可以利用基站直接提供

的计算和存储服务来提高用户的体验.首先,将这些视频数据传递给基站内部署的近端云服务器.在这些服务器上,可以对采集的视频数据进行合成和处理,形成不同高低分辨率的视频数据,让不同的用户进行选择 and 观看.

同时,利用近端云服务器提供的存储能力,热门视频可以被缓存在这些服务器上供终端用户使用,这不仅会减少用户的下载延迟,也会减少主干网的传输压力.与现有优化视频分享的 CDN(Content Delivery Network)技术需要部署专门服务器然而却离用户还有几跳距离相比,利用 5G 基站中缓存的视频分发加速效果将更加显著.

再以提高智能移动终端计算能力的代码或计算卸载为例.由于智能移动终端和可穿戴式设备在计算和存储资源等方面的能力有限,因此如何利用云计算的强大功能来帮助完成自身所需要的计算或存储任务,是一条解决智能移动终端资源受限问题的良好途径.然而,由于这些卸载的用户体验严重依赖于网络延迟,因此这些卸载任务不太适合部署到远端云计算中心来执行^[20].但是利用近端云计算设施来实现代码卸载,不仅可以提高执行效率,还能节约智能终端的能耗.在 4G LTE 网络下,把智能手机上的 Tesseract-OCR 代码迁移到 Cloudlet 平台执行,与在 Amazon EC2-West 上执行相比,不但速度得到提高,而且能耗降低一半多^[35].显而易见,能耗的降低对于智能移动终端而言是非常具有吸引力的一个特性.

5G 网络将支持 D2D 通信,即两个终端设备可以不经基站的转发而直接进行设备间通信,以降低基站的通信处理压力.利用这一特性, FemtoClouds^[36]可以在多个终端设备间通过相互协作而组成一个动态和自我配置的“毫微云”. FemtoClouds 利用附近未使用的终端设备在网络边缘提供计算服务,从而减少了把计算卸载到传统云数据中心带来的网络延迟.在 FemtoClouds 中,终端设备通过网络加入终端集群,并将设备信息(包括计算可用性、可共享的计算、使用历史等)和共享策略发送到集群中,进而可以向其它终端发送计算代码和输入数据来卸载计算密集型的任务.

5.3 智能电网

智能电网(Smart Grid)是一个自动化、智能化的电力输送网络,它保证了电力从发电厂产生直到被用户电器设备消耗的整个电力传输过程中电力的智能化监控和管理^[37].智能电网集成了分布式系统、

网络通信技术以及自动控制理论等来保证电能市场交易的实时进行和电网上各成员的无缝连接及实时互动.智能电网有三个层面的含义:首先是通过传感设备,对发电、输电、配电和供电等过程中的关键设备进行实时监控;然后将监控获得的电网数据通过网络系统进行收集并整合;最后通过对监控数据的挖掘分析,实现对整个电力系统的优化.

近端云计算可以很容易地解决智能电网发展过程中遇到的问题并使其获益.相对于现有的集中式的以及完全分布式的智能电网模型,利用近端云计算模式,智能电网中集中的数据采集与监控系统(Supervisory Control And Data Acquisition, SCADA)可以通过分布式的智能电表(Smart Meters)和微电网(Micro-grids)进行补充,这不仅可以改善智能电网的可扩展性、成本、安全性以及电网系统的响应时间,还可以将一些可再生能源和分布式能源(例如,风电场、太阳能光伏电场等)接入到主干电网^[15].

在近端云计算模式中,智能电网将变成一个层次之间相互作用的层次化系统^[2,38].第一层是电网传感层,负责收集数据、处理一些要求实时响应的数据以及触发执行器的控制命令.同时,它还会过滤要在本地进行处理的数据,并将其余的数据发送到更高层;第二层是微电网层,负责数据的可视化和报告(人与机器之间的交互)、系统之间的交互以及数据的处理(机器与机器之间的交互).第三层是 SCADA 和远端云计算中心层,负责长时间的数据存储和商业智能分析.在这样一个层次化的系统中,层次越高意味着地理覆盖面积越大,相互交互的时间延迟越长.因为层次之间交互的时间可能从秒到分钟(实时分析),甚至数天(事务分析),这要求利用近端云计算的智能电网系统需要在低层支持数据的临时存储,在高层支持数据的半永久存储.

5.4 虚拟/增强现实

近年来,随着计算机视觉、人工智能、大数据分析以及计算和存储技术的发展,虚拟现实、增强现实以及混合现实等沉浸式技术得到了快速发展和应用.这些技术可以实时地将一个世界的不同视觉景象、声音以及情感以高分辨率的声光传递并转换到另一个世界.虽然现阶段虚拟现实技术主要应用在游戏、商业会议以及残疾人辅助等方面,但是可以预见在不远的将来,虚拟现实或者混合现实等技术将会应用到我们日常生活的多个方面,例如远程聊天、社交活动等等.

然而,目前的虚拟现实技术和系统还不能进行

大量复杂的计算和数据处理,例如,捕获和处理用户面部的表情和感觉细节.同时,虚拟世界的呈现需要传输大量的视频帧数据,因此只能利用高速的有线网络进行视频数据传输.无线网络的低带宽和高延迟还难以满足虚拟现实系统的需要,这在一定程度上限制了虚拟现实应用的广度和深度^[39].

近端云计算为解决虚拟现实等沉浸式技术的应用障碍提供了一条非常有用的途径.首先,近端云基础设施能弥补虚拟现实设备的计算和存储能力的不足.虚拟现实设备可以将用户面部表情细节处理等需要大计算量的任务交由近端云服务器来完成.其次,通过在近端云服务器上缓存与远程用户互动过程中产生的视频和其它数据,可以减少网络数据的传输量并提高响应速度.同时,利用近端云基础设施提供的短距离的终端到终端的超高无线近距离传输通道,虚拟现实系统将不但可以摆脱有线网络的限制,而且可以利用无处不在的无线网络技术,实现虚拟现实设备和系统间的无缝连接,从而实现虚拟现实和物理现实的融合.

文献[40]提出了一种利用相邻智能移动终端的 GPU 来帮助进行图像渲染的系统,通过把渲染任务交付给相邻智能设备来执行,不但可以实现高分辨率的图像渲染效果,提高用户体验,而且可以节省能耗,提高终端的续航能力.这在一定程度上说明虽然在虚拟/增强现实领域利用近端云计算还需要克服一些具体的技术障碍,但是其应用前景非常广阔.

5.5 无人机

无人驾驶飞机(Unmanned Aerial Vehicle),简称无人机,是指利用无线电遥控设备和自有程序控制装置操控的不载人飞机^[41].无人机在军事和民用领域有非常广泛的应用.在军事领域,无人机可以作为侦察机和靶机等使用.在民用领域,无人机的应用更加广泛,包括航拍、灾难救援、快递运输、测绘、新闻报道、野生动物保护等.基于其广阔的应用前景,无人机的应用开发正在如火如荼的进行中.例如,在快递领域,无人机可以实现精准及时的快递配送,有效地节省资源.因此,国内的顺丰以及国外的谷歌、亚马逊等公司都在研发测试自己的无人机快递服务.

尽管无人机的应用前景非常广阔,但其并未得到大规模的应用.这主要是由于无人机在技术、安全和监管方面面临着诸多尚未解决的问题.在技术方面,无人机对于网络延迟的要求比较高,而基于云计算的处理数据模式延迟比较大,因而很难满足其要

求.例如,无人机在进入停机位的过程中,一些无人机的速度非常高(可以达到 10 英里每小时).所有消息通过集中的云计算中心处理将产生较大的延迟,等到无人机报告问题时,可能已经错过处理的最佳时间.其次,无人机面临着严峻的安全问题^[42].同一空域可能同时出现执行不同任务的不同公司的无人机,而且还可能存在飞鸟以及高层建筑等,如何避免无人机与无人机以及与其它空中物体之间的相互碰撞,是需要重点解决的安全问题.最后,无人机还面临着复杂的监管问题^[43].在一些敏感区域或者民航飞机航道上,各地的航空管理局是禁止无人机飞行的.因此,无人机需要在复杂的监管环境中进行飞行.如何根据地理位置有效地控制无人机的飞行是监管无人机首要解决的一个问题.

针对无人机应用面临的挑战,近端云计算给出了可行的解决方案.首先,近端云计算可以满足无人机低延迟的需求.根据任务的延迟需求,近端云计算可以在不同的层次进行数据处理,满足不同延迟的消息处理需求.例如,将一部分无人机收集的视频数据的处理任务卸载到边缘云数据中心上,可以很大程度上减小数据处理的延迟^[44].其次,近端云计算可以从无人机自身、相邻无人机、近端云计算设施(例如,在无人机航线或高层建筑物上部署的节点)、远端云计算中心多个层次来保证无人机的安全.最后,近端云计算可以实现基于地理位置的无人机监管服务.对于禁飞区域,可以在近端云计算设施进行配置,当无人机接近禁飞的近端云计算设施时,实现主动避让.另外,还可以在近端云计算设施和远端云计算中心对无人机飞行的状况和轨迹进行监控,通过对区域内无人机的信息进行集中地整理,监管部门

可以有效地管理区域内的无人机.

反过来,无人机也可以作为一种实现近端云计算的平台或技术.例如,无人机可以为终端设备提供通信中继服务^[45].在需要临时增加网络覆盖或者发生过自然灾害后通信被破坏的地区,无人机可以作为飞行的基站与卫星或其它通信基站进行通信来保证所在区域内的通信服务.无人机还可以作为一个移动的朵云平台^[46],为其它的无线终端设备提供一个小型的云计算平台,使得计算等任务可以从终端设备卸载到无人机上.在无线传感器网络中,无人机可以作为一个移动的数据收集平台,将各类传感器采集的数据统一地收集起来^[44].因此,相信随着无人机与近端云计算技术的不断发展,更多实际的应用问题可以通过无人机与近端云计算技术的相结合而得到有效解决.

6 近端云计算面临的技术挑战

尽管近端云计算拥有广阔的应用前景和发展机遇,能够解决云计算所面临的难题,但是实现和部署近端云计算也面临着诸多挑战,包括商业挑战和技术挑战等.商业挑战主要由公司来推动和解决,本文不做分析.本文主要分析近端云计算面临的技术挑战.理清技术挑战有助于知道哪些方面是实现近端云计算模式需要重点解决的问题,从而为以后近端云计算的研究指明方向.

如表 4 所示,本节将从体系结构和模型、网络接入和管理、资源管理和调度、定价和激励机制、安全和隐私保护等方面分别介绍近端云计算面临的技术挑战.

表 4 近端云计算面临的技术挑战

技术挑战	主要问题	相关研究	
体系结构和模型	体系结构	在分布式、异构、能力各异的资源上构建一个可实施部署的体系结构	[2] [3] [18] [47] [48] [49]
	编程模型	建立一个可以根据连接方式和移动特征等选择合适的计算节点的编程模型	[50] [51] [52] [53] [54] [55] [56]
网络接入和管理	/	管理一个异构、复杂、多变的网络,在保证网络的联通性的同时提供服务	[57] [58] [59] [60] [61] [62]
资源管理和调度	资源隔离	屏蔽各类设备在操作系统和硬件上的差异,并实现应用程序的隔离	[20] [56] [63] [64] [65]
	资源发现共享和配置	综合考虑设备的移动、服务的可用性和收益等因素的资源发现、共享和配置	[66] [67] [68] [69]
	资源调度	任务能够动态地适应端设备、网络、资源等因素的变化在不同的地点执行	[70] [71] [72] [73] [74] [75]
定价与激励机制	/	动态的按需定价和有效的参与激励机制	[76] [77] [78] [79] [80]
安全和隐私保护	身份认证	不依赖于第三方的、基于地理位置的、轻量化的认证技术	[81] [82] [83] [84]
	访问控制	支持多种用户决策、基于地理位置、满足不同目标和限制的访问控制机制	[85] [86] [87]
	隐私保护	轻量级的敏感数据隐私、使用隐私和位置隐私保护机制	[88] [89] [90] [91]

6.1 体系结构和模型

体系结构和模型是实现及应用近端云计算模式的基础.因此,首先需要研究近端云计算的体系结构和模型.

6.1.1 体系结构

相对于集中部署的云计算来说,近端云计算的体系结构更加复杂.首先,近端云计算中资源的种类更多,既包括集中部署在数据中心中的资源,也包括分布在网络上的资源(例如,路由器、边缘服务器等),以及动态移动的终端资源等.其次,资源是分布式和小型集中式相结合的,这些资源的地理位置分布范围很广.资源之间相互连接的方式不仅包括有线,还包括各种无线连接方式(例如,Wi-Fi、Wi-Max、蓝牙、3G、LTE等).然而,这些连接方式的传输速率和延迟各异.最后,近端云计算模式中各类资源往往是异构的,而且能力差异大.如何在这些分布式、异构、能力各异的资源上构建一个可实施部署的体系结构非常具有挑战性.

现有的针对雾计算^[2]和露计算^[18]的体系结构研究工作,只是给出了简单的层次划分,并没有给出可实施部署的体系结构.虽然针对移动边缘计算的体系结构研究相对较多^[3,47,48],但与具有普遍意义的近端云计算相比,移动边缘计算没有考虑终端设备的协作.

近端云计算的体系结构研究,可以参考现有的云计算体系结构^[49]和移动边缘计算的体系结构^[47],将原有的体系结构向近端设备(例如,路由器、边缘服务器)和终端设备进行扩展和扩充,并考虑终端设备之间的相互协作.针对近端云计算体系结构的研究,可以先设计一个较为通用的体系结构模型标准.在此基础上,再针对特定的应用场景研究适合不同应用场景需求的体系结构.

6.1.2 编程模型

对于近端云计算来说,研究类似于云计算中的MapReduce^[50]编程模型是非常必须的.因为统一的编程模型可以隐藏近端云计算模式中的复杂管理功能,让开发人员不用考虑近端云计算模式中动态、异构、层次化的各类资源,只需要关注应用本身的逻辑,从而加快近端云计算模式应用的开发.但是,和云计算相比,近端云计算的编程模型更加的复杂.首先,计算资源的种类繁多,而计算可以在任何资源节点上进行,如何协调同一计算任务的执行地点,并汇总计算结果是一项具有挑战性的任务;其次,近端云计算中节点的连接方式各异,而且存在移动的节点.因

此,在任务分发之后,一些计算节点在返回结果之前可能已经无法访问.如何根据节点的连接方式和移动特征,选择合适的参与节点也非常具有挑战性.

Hong等^[51]针对未来互联网应用提出了一种高级编程模型,该模型具有可按需扩展以及大规模地理位置分布和延迟敏感的特点.然而,其中的计算节点是固定的.Orsini等^[52]提出的CloudAware通过一个上下文管理器来估计网络的连接状态,并通过求解优化问题(例如,最小化计算时间)来切分应用和执行卸载策略.不过,CloudAware只是一个抽象的编程模型.Foglets^[53]支持发现与应用程序需求匹配的资源,并可以根据传感器的移动特征和动态的计算需求对应用程序的组件进行迁移.MCEP^[54]通过实时迁移,能够根据位置信息按需动态地配置计算资源,从而降低延迟、网络利用率和处理开销.但是Foglets和MCEP都是基于自己定义的数据模型和应用程序接口.FogFlow^[55]利用标准化的应用程序接口来实现云和边缘之间的编程模型.不过,目前FogFlow只针对智慧城市,如果应用到其它的场景中,还需要进行适当的修改.IOx¹提供了一个可以无缝跨越不同网络边缘设备的应用程序支持框架和计算平台.IOx通过开放式的应用程序开发环境向开发者提供了对应用程序整个生命周期(包括:开发、分发、部署、托管、监视和管理)的管控.但是IOx并不开放而且依赖于很昂贵的硬件.ParaDrop^[56]是一种边缘计算框架,它可以使第三方开发者利用距离终端最近的接入设备(例如,Wi-Fi接入点或家用机顶盒等)上的计算资源.ParaDrop通过开发者API支持服务的动态安装,并且支持基于策略的资源管理.但是ParaDrop主要针对轻量化的应用程序.

为了加快近端云计算模式的应用,需要研究结合近端云计算模式特征的通用编程模型.这种通用编程模型可以根据计算任务的需求(例如,延迟要求等),能够在端、相邻端、近端云计算设施和远端云计算中心中选择一个或者多个节点执行计算任务.具体在哪里执行,开发人员可以自己定义,也可以根据一定的匹配算法或者机器学习的方法,综合考虑任务的需求让任务自动地选择执行节点.

6.2 网络接入和管理

和传统云计算相比,近端云计算的网络更加复杂多变.首先,近端云计算的网络往往是异构的,既包

¹ IOx - App Developer's guide, <https://developer.cisco.com/media/iox-dev-guide-3-10-16/> 2017,7,10

括稳定可靠的有线连接,也包括存在不确定性的无线连接;其次,近端云计算的网络由于端设备的移动,使得端设备连接的网络会随着位置的变化而变化,而且端设备连接的网络存在断开的可能;再者,近端云计算支持端设备之间的相互协作,这要求端设备可以在不依赖于其它网络基础设施的情况下构建协作网络,而且网络需要支持节点的加入、离开和移动.管理这样一个异构、复杂、多变的网络,保证网络的联通性,并在其上提供服务非常具有挑战性.

最近一些新兴的网络技术,例如软件定义网络(SDN)和网络功能虚拟化(NFV),为创建灵活且易于管理的网络提供了解决方案.将SDN和NFV技术应用于近端云计算模式可以简化网络实施和管理的复杂度,增加网络的可扩展性并降低成本.ETSI ISG MEC建议通过SDN和NFV管理网络资源,不仅可以减少部署开销,还可以提供一个通用的业务流程管理基础架构^[57].在近端云计算中,集成SDN技术最大的挑战是如何适应终端设备的移动性以及不可靠的无线网络连接.Ku等^[58]提出了用于移动/车载自组织网络(MANET/VANET)的基于SDN的移动云架构,其基于SDN的频率选择架构可以适应异构的无线网络接口,并支持特权业务预留和调频通信等应用.除此之外,如何放置不同的网络控制器(包括持久连接的边缘设备和间歇连接的终端设备)以及在哪里部署网络控制器也是在近端云计算中应用SDN需要解决的问题^[59].解决这些问题需要根据具体的近端云计算应用的需求,研究针对性的SDN架构和系统.

NFV可以通过虚拟化近端云计算中的网关、交换机、防火墙等设备来增强近端云计算中网络资源管理的灵活性.NFV已经在数据中心网络^[60]、蜂窝核心网络^[61]中得到了应用.相比之下,这些网络中的设备具有丰富的计算资源,而近端云计算中很多终端设备的资源非常的有限.因此,在近端云计算应用NFV需要重点考虑有限的资源问题和虚拟化网络设备性能^[62],同时也需要结合近端云计算应用和服务的特定需求.

6.3 资源管理和调度

资源管理和调度是各类应用程序在近端云计算系统中运行的基础.本节从资源隔离、资源发现、共享和配置、资源调度三个方面来介绍近端云计算资源管理和调度面临的挑战.

6.3.1 资源隔离

为了利用各类计算设备的资源,需要解决资源

的隔离问题.首先,因为不同计算设备的软硬件是异构的,要想把这些异构的资源利用起来,就需要屏蔽设备在操作系统和硬件上的差异;其次,同一计算设备上往往会运行多个应用,这些应用会共享设备的资源.为了避免因为一个应用程序的失败而造成其它应用或者整个系统的崩溃,就需要将这些应用隔离开来;最后,在设备上,特别是在终端设备上,往往存储着用户的隐私数据.为了保护用户的隐私和安全,往往需要隔离第三方的应用程序.

虚拟化技术是实现资源隔离的一种有效手段.常用的虚拟化技术包括虚拟机监控器(Hypervisor)和容器(Container).虚拟机监控器可以实现系统级的虚拟化,是云计算中常用的虚拟化技术.Cloudlets^[20,63]就利用了这种虚拟化技术.相对于虚拟机监控器来说,容器是更轻量级的虚拟化技术.ParaDrop^[56]则采用了容器虚拟化技术.因为容器是在操作系统层实现的虚拟化技术,因此相对于虚拟机监控器来说,其灵活性更低.Condroid^[64]是一种基于容器的轻量级虚拟化技术,使得一个Android设备可以托管多个独立的Android虚拟机.但即便如此,虚拟机监控器和容器这些系统级的虚拟化技术仍然无法支持一些资源更少的设备(例如,IoT等)的资源隔离需求.FlowFence^[65]通过构建一些隔离模块并在sandbox中操作敏感数据,从而可以实现对IoT应用中敏感数据的有效控制.但是,支持FlowFence需要按照FlowFence的逻辑重写相应的IoT应用.

因为近端云计算中设备的能力差异很大,具体采用哪种资源隔离技术,还需要根据设备的计算能力和应用场景的需求进行综合考量.此外,访问控制可以通过隔离第三方应用可访问的内容,从而有效地解决资源隔离的问题.

6.3.2 资源发现、共享和配置

资源发现和共享(Resource Discovery and Sharing)是终端设备实现对等协作以及利用近端云计算设施中资源的基础.因为终端设备的位置会变化,在位置变化之后,终端设备需要能够及时发现附近可用的其它终端设备以及近端云计算设施.而且,终端设备也需要能够在新的位置将自身空闲的资源共享给其它设备使用.Liu等^[66]提出了一种自适应的资源发现方法,该方法根据网络环境的不同自动地在集中式和洪泛策略之间进行切换,以节省能耗.Nishio等^[67]提出了一种异构资源共享框架,通过将诸如CPU、通信带宽和存储等异构资源映射成“时间”资源.因此,资源共享问题可以形式化成一个

最大化面向服务的效用函数问题.不过,这里的效用函数只考虑了延迟因素.然而,在近端云计算中,资源发现和共享除了需要考虑能耗和延迟因素外,还需要考虑设备的移动、服务的可用性甚至收益等因素.

资源配置(Resource Provisioning)是在用户请求资源时向用户分配其所需的资源.Aazam 等^[68]提出了一种基于资源需求预测的资源配置方法.通过综合考虑用户的需求、服务的类型和价格模型定义支出函数,从而提前为用户分配资源.然而,对于很多应用来说,很难准确预测其需求的资源量.Skarlat 等^[69]提出了一个雾计算资源配置的概念框架,通过形式化一个优化问题在运行时进行资源配置.但是,由于优化模型所考虑因素的限制,使得这个框架的通用性非常有限.近端云计算的资源配置需要综合考虑应用的需求和资源的存储、计算、网络带宽以及价格等因素.除此之外,近端云计算的资源配置需要满足应用延迟的要求,因此还需要考虑延迟因素.此外,近端云计算的资源配置还需要考虑设备的移动性.因为终端设备的移动会造成可用资源的状态和网络延迟产生变化.

6.3.3 资源调度

近端云计算中的资源调度之所以具有挑战性,一方面是因为近端云计算中计算、存储等功能可以在端设备本身、相邻端设备、近端云计算设施和远端云计算中心中的任何一处执行,可供选择的执行位置较多;另一方面是因为近端云计算中资源调度除了需要考虑计算或存储等任务完成所需要的资源因素外,还需要考虑终端的移动性、延迟等因素.要解决近端云计算中的资源调度问题需要解决这些问题:如何实现端与端之间的对等协作,以及终端和近端云计算设施、远端云计算中心之间的资源调度;在不同的层次之间选择什么样的调度粒度;如何切分应用程序,使其可以在端设备、近端云计算设施和远端云计算中心之间实现并发和同步;如何执行调度决策,使得调度能够动态地适应端设备、网络、资源等因素的变化.

弹性资源调度是云计算最主要的特征之一.因此,作为云计算向终端设备的扩展,近端云计算也应当能够支持资源的弹性调度.Zhan 等^[70]总结了云计算中弹性资源调度方法.Kavvadia 等^[71]提出了一种基于无容量限制设施选址的弹性虚拟机放置方法.Liu 等^[72]提出了一种激进式的资源调度方法,通过激进的资源分配策略来保证服务质量.这些弹性资源调度方法大都以虚拟机为粒度进行资源调度.

相比之下,近端云计算则需要考虑更加细粒度的弹性资源调度,而且在资源调度的过程中需要结合设备的位置信息、延迟要求等多种因素.

计算卸载(Computation Offloading)也是资源调度的一种形式,通过将一部分任务迁移到其它设备执行来扩展终端设备的能力,从而有效地解决终端设备资源有限问题.Chun 等^[73]提出了应用程序在端设备和云之间迁移的方法 CloneCloud,它使用静态代码分析器在应用程序字节码中自动标记可能迁移的点.Barbera 等^[74]研究了移动终端利用云平台进行计算卸载以降低移动终端的带宽和能量消耗的方法.Kosta 等^[75]提出了一个卸载手机应用到云平台的计算架构 ThinkAir.这些计算卸载方法也可以应用到近端云计算的计算卸载中.和上述现有的计算卸载方法相比,近端云计算中的计算卸载需要考虑计算执行的位置、端设备的移动性和延迟需求等多种因素,以使得应用和服务能够通过计算卸载以充分利用近端云计算的优势.

6.4 定价和激励机制

从已有近端云计算的相关研究工作可以知道,近端云计算的参与者包括:1)云服务提供商,希望将自己的服务扩展到终端设备上;2)互联网服务提供商和无线网络运营商,可以构建近端服务设施来连接云服务提供商和终端用户;3)终端用户,可以共享自己空闲的计算、存储和网络等资源,也可以通过利用其它终端或者服务提供商提供的资源以扩充自身设备的能力.为了保证各个参与者的利益,实现类似于云计算的“随用随付”(pay-as-you-go)的原则,首先需要解决的就是定价问题.例如,如何为不同参与者的资源进行定价,以及如何在不同的参与者之间进行收益的分配等^[76].另外,还可以参考云计算中的动态定价机制^[77],动态地给近端云计算的资源服务进行定价,以最大限度地提高收益和资源利用率.再者,近端云计算支持端设备之间组件自组织的协作网络,因此还需要结合近端云计算应用的特点研究不依赖于第三方或者基站协助的定价机制^[78].

激励机制(Incentive Mechanism)可以激励终端用户参与到近端云计算系统中.因此,为了加速近端云计算的实现和部署,需要研究有效的激励机制.但是,因为终端用户的存在,使得近端云计算中影响参与者参与的因素不再是单纯的经济收益.例如,如果终端用户共享自己的资源可以帮助自己的亲人或者朋友,终端用户参与的积极性会提高^[79].因此,近端云计算的激励机制可以结合用户的社交属性来激

励更多的用户参与到资源的共享合作中来.再者,如果终端用户的参与可以提高自身的计算体验的话,终端用户也会非常乐意参与到近端云计算中^[80].因此,在近端云计算的激励机制研究中要充分考虑各个参与者的特点,通过挖掘各个参与者的参与动机从而激励用户参与到近端云计算中.另外,还可以结合应用场景的特点,将激励转换成虚拟货币等较为通用的激励,使得用户可以利用通用的虚拟货币来换取自身所需的资源.

6.5 安全和隐私保护

随着人们安全和隐私保护意识的增强,安全和隐私保护变得越来越重要.与云计算相比,近端云计算既具有在本地或近端解决安全和隐私保护问题的优势,也同时面临着更加严峻的安全和隐私保护问题^[92].首先,近端云计算涵盖了很多终端设备,而这些终端设备存在潜在的安全问题.以 Android 设备为例来说,有 82% 的 Android 设备存在 25 个安全漏洞中的至少一个^[93];其次,近端云计算中终端设备接入的网络多种多样,而这些网络的安全性难以保障,使其更容易受到攻击.以 Wi-Fi 网络为例,据统计¹,49% 的家用 Wi-Fi 是不安全的,80% 的家用路由器使用的是默认密码;89% 的公共 Wi-Fi 热点是不安全的.最后,在近端云计算中,节点的分布横跨了端、网络(近端)和云(远端),跨域实施安全和隐私保护的难度更大.而且,一些传感设备的资源非常有限,这就造成现有的很多安全保护技术无法被直接使用,因为它们所需要的资源无法得到满足^[94].下面讨论一些可以用于实现近端云计算安全和隐私保护的技术.

6.5.1 身份认证

身份认证是安全保护的第一道防线.近端云计算中的身份认证面临着如下挑战.首先,近端云计算支持端设备的相互协作,因此需要支持不依赖第三方认证设施的身份认证技术;其次,近端云计算中很多终端设备的资源是有限的,因此轻量级的身份认证技术是非常必要的;再次,近端云计算支持地理位置的服务,如何实现基于地理位置的身份认证,以及实现在用户位置切换之后,不需要用户重新进行身份认证也是一个问题;最后,如何在近端云计算的多个架构层次之间完成身份认证.已有的身份认证技术很多,部分可以应用到近端云计算中来克服上述挑战.Toosi 等^[81]对云计算中的身份认证技术进行了

总结.Ibrahim 等^[82]提出了一种身份认证系统 Octopus,允许任何雾计算用户和雾计算节点相互认证,但是这种方法中的雾计算节点存储了信任域内所有用户的凭据信息.Bouzefrane 等^[83]使用近场通信(Near Field Communication, NFC)来验证终端设备是否能将任务卸载到已认证的本地云中.Yang 等^[84]则提出了一种允许终端用户从一个区域迁移到另一个区域的身份认证技术.除此之外,还可以借鉴云计算和 P2P 网络中的身份认证技术,通过在这些现有的身份认证技术中增加位置等信息来使得其更加适合于近端云计算的应用场景.例如,在认证过程中增加诸如时间、所在位置、设备等状态信息,不仅可以增强认证的安全性,还可以简化某些情况下的认证过程.

6.5.2 访问控制

访问控制在身份认证的基础上,根据身份的合法性对资源访问请求加以控制.访问控制技术已经成为了云计算可靠的安全保障措施之一^[85].Almutairi 等^[86]提出了一种应用于多云环境的基于角色控制策略的分布式访问控制框架,以支持域角色映射和限制验证.Dsouza 等^[87]提出了可以用于雾计算的基于策略的资源访问控制方法,以支持异构资源之间的安全协作和互操作.在近端云计算中应用访问控制技术还需要解决下面这些问题:如何明确近端云计算中的用户角色以及权限范围;如何结合近端云计算的地理位置特征,实现基于地理位置的访问控制;如何在不同的层次中分发权限控制策略以满足不同层的目标和资源限制需求等.解决以上这些问题需要结合具体的近端云计算应用场景的特点,针对性地设计访问控制策略.

6.5.3 隐私保护

和云计算相比,近端云计算由于覆盖了各种各样的用户终端设备,使得应用可以更容易地收集用户的敏感数据.另外,端设备的资源往往非常有限,而隐私保护算法往往都需要消耗很多的资源,这就需要将很多用户的敏感数据传输到近端或者远端云计算中心进行处理.如果在网络上明文传输敏感数据,会造成用户的隐私泄露.同态加密(Homomorphic Encryption)技术可以实现在不解密的情况下对敏感数据进行处理,因此可以实现在近端或者远端云计算中心处理敏感数据的同时保护用户的隐私^[88].另外,差分隐私(Differential Privacy)可以用于确保在统计查询的情况下不泄露数据集中任意单个条目的隐私^[89].除了数据隐私外,使用隐私和位置隐私也是

¹ Wifi Network Security Statistics/Graph, <http://graphs.net/wifi-stats.html> 2017,7,10

近端云计算隐私保护的重要内容^[90].例如,因为终端设备使用附近近端云计算设施中的资源,使得近端云计算设施可以很容易地根据自身的位置,以及用户的使用时间和使用模式等信息来推测出用户的地理位置信息和使用隐私.使用隐私和位置隐私最有效的保护方式是匿名和身份混淆,即便近端云计算设施知道用户在附近,但也无法识别出该用户^[91].另外,用户在选择近端云计算设施时,可以不单纯地根据位置信息进行决策,而是结合延迟、用户评价等多种因素动态地选择近端云计算设施,以避免近端云计算设施通过其本身的位置推测出用户的位置.

7 结论

本文讨论了云计算在发展过程中所遇到的瓶颈,揭示了云计算的基本特征及其发展演化的必然趋势.在此基础上,对后云计算时代出现的新型网络计算模式进行了简单介绍,并总结了这些网络计算模式的基本特征.接着,对几种近端云计算模式,即雾计算、移动边缘计算和边缘计算等进行了详细地分析、比较和总结,使读者能够掌握它们各自的特点以及细微区别.另外,为了使读者能够清晰地了解近端云计算的发展前景及其面临的困难,本文还通过一些具体的应用例子分析和总结了近端云计算发展所面临的机遇和技术挑战.相信随着云计算问题的日益突出以及近端云计算实践经验的不断积累,在不久的将来,近端云计算无疑会成为后云计算时代产业界和学术界开发和研究的一个新热点.

参考文献

- [1] Bonomi F. Connected vehicles, the Internet of Things, and fog computing//Keynotes of the eighth ACM international workshop on vehicular inter-networking. Las Vegas, USA, 2011
- [2] Bonomi F, Milito R, Zhu J, Addepalli S. Fog computing and its role in the internet of things//Proceedings of the first edition of the MCC workshop on Mobile cloud computing. Helsinki, Finland, 2012: 13-16
- [3] Patel M, Hu Y, Chan C, et al. Mobile-edge computing. France: European Telecommunications Standards Institute, Introductory Technical White Paper, 2014
- [4] Satyanarayanan M. The emergence of edge computing. IEEE Computer, 2017, 50(1): 30-39
- [5] Berman F, Fox G, and Hey AJG. Grid Computing: Making the Global Infrastructure a Reality. West Sussex, England: John Wiley, 2003
- [6] Papazoglou MP. Service-oriented computing: Concepts, characteristics and directions//Proceedings of the fourth international conference on web information systems engineering. Roma, Italy, 2003: 3-12
- [7] Murch R. Autonomic Computing. USA: IBM Press, 2004
- [8] Kephart JO, Chess DM. The vision of autonomic computing. Computer, 2003, 36(1): 41-50
- [9] Zhang Y, Zhou Y. Transparent computing: a new paradigm for pervasive computing//Proceedings of the 3rd international conference on ubiquitous intelligence and computing. Wuhan, China, 2006: 1-11
- [10] Grigorik I. High Performance Browser Networking. Sebastopol, California, USA: O'Reilly, 2013
- [11] Tolia N, Andersen DG, Satyanarayanan M. Quantifying interactive user experience on thin clients. Computer, 2006, 39(3): 46-52
- [12] Stojmenovic I, Wen S. The fog computing paradigm: scenarios and security issues//Proceedings of the 2014 federated conference on computer science and information systems. Warsaw, Poland, 2014: 1-8
- [13] Luan TH, Gao L, Li Z, Xiang Y, Wei G, and Sun L. Fog computing: focusing on mobile users at the edge. Australia: Deakin University, Preprint: arXiv:1502.01815, 2016
- [14] Vaquero LM, Rodero-Merino L. Finding your way in the fog: towards a comprehensive definition of fog computing. ACM SIGCOMM Computer Communication Review, 2014, 44(5): 27-32
- [15] Yi S, Hao Z, Qin Z, Li Q. Fog computing: Platform and applications//Proceedings of the Third IEEE workshop on hot topics in web systems and technologies. Washington, USA, 2015: 73-78
- [16] Yi S, Li C, Li Q. A survey of fog computing: concepts, applications and issues//Proceedings of the 2015 workshop on mobile big data. Hangzhou, China, 2015: 37-42
- [17] Ahmed A, Ahmed E. A survey on mobile edge computing//Proceedings of the 10th international conference on intelligent systems and control. Coimbatore, Tamilnadu, India, 2016: 1-8
- [18] Wang Y. Cloud-dew architecture. International Journal of Cloud Computing, 2015, 4(3): 199-210
- [19] Ha K. System Infrastructure for Mobile-Cloud Convergence [Ph.D. Thesis]. Carnegie Mellon University, Pittsburgh, PA, USA, 2016
- [20] Satyanarayanan M, Bahl P, Caceres R, Davies N. The case for vm-based cloudlets in mobile computing. IEEE Pervasive Computing, 2009, 8(4): 14-23
- [21] Wang Y. Definition and categorization of dew computing. Open Journal of Cloud Computing, 2016, 3(1): 1-7
- [22] Skala K, Davidovic D, Afgan E, Sovic I, Sojat Z. Scalable distributed computing hierarchy: Cloud, fog and dew computing. Open Journal of Cloud Computing, 2015, 2(1): 16-24
- [23] Zhou Y, Tang W, Zhang D, Zhang Y. Software-defined streaming-based code scheduling for transparent computing//Proceedings of the fourth international conference on advanced cloud and big data. Chengdu, China, 2016: 296-303
- [24] Zhang Y, Guo K, Ren J, Zhou Y, Wang J, Chen J. Transparent computing: A promising network computing paradigm, Computing in Science & Engineering, 2017, 19(1): 7-20
- [25] Sun QB, Liu J, Li S, Fan CX, Sun JJ. Internet of Things: Summarize on

- concepts, architecture and key technology problem. *Journal of Beijing University of Posts and Telecommunications*, 2010, 33(3): 1-9 (in Chinese)
- (孙其博, 刘杰, 黎彝, 范春晓, 孙娟娟. 物联网: 概念、架构与关键技术研究综述. *北京邮电大学学报*, 2010, 33(3): 1-9)
- [26] ITU. *The Internet of Things*. Geneva: International Telecommunication Union, ITU Internet Reports, 2005
- [27] Cisco. *Fog computing and the Internet of Things: extend the cloud to where the things are*. San Jose, California, USA: Cisco, White paper, 2015
- [28] Perera C, Qin Y, Estrella JC, Reiff-Marganiec S, Vasilakos AV. Fog computing for sustainable smart cities: a survey. *ACM Computing Surveys*, 2017, 50(3): 1-43
- [29] Datta K. S, Bonnet C, and Haeri J. Fog Computing architecture to enable consumer centric Internet of Things services//*Proceedings of the 2015 international symposium on consumer electronics*. Madrid, Spain, 2015: 1-2
- [30] Tao M, Ota K, Dong M. Foud: Integrating fog and cloud for 5G-enabled V2G networks. *IEEE Network*, 2017, 31(2): 8-13
- [31] Ren J, Guo H, Xu C, Zhang Y. Serving at the edge: A scalable IoT architecture based on transparent computing. *IEEE Network*, 2017, 31(5): 96-105
- [32] Jalali F, Smith OJ, Lynar T, Suits F. Cognitive IoT gateways: Automatic task sharing and switching between cloud and edge/fog computing//*Proceedings of the SIGCOMM posters and demos*. New York, USA, 2017: 121-123
- [33] 3GPP. Technical specification group services and system aspects; system architecture for the 5G systems; stage 2 (release 15). 3GPP TS 23.501 V0.4.0, 2017
- [34] Anjum A, Abdullah T, Tariq M, Baltaci Y, Antonopoulos N. Video stream analysis in clouds: An object detection and classification framework for high performance video analytics. *IEEE Transactions on Cloud Computing*, 2016, PP(99): 1-17
- [35] Hu W, Gao Y, Ha K, Wang J, Amos B, Chen Z, Pillai P, Satyanarayanan M. Quantifying the impact of edge computing on mobile applications//*Proceedings of the 7th ACM SIGOPS Asia-Pacific workshop on systems*. Hong Kong, China, 2016: 1-8
- [36] Habak K, Ammar M, Harras KA, and Zegura E. FemtoClouds: leveraging mobile devices to provide cloud service at the edge//*Proceedings of the 8th international conference on cloud computing*. New York, USA, 2015: 9-16
- [37] Chen SY, Song SF, Li LX, Shen J. Survey on smart grid technology. *Power System Technology*, 2009, 33(8): 1-7 (in Chinese)
- (陈树勇, 宋书芳, 李兰欣, 沈杰. 智能电网技术综述. *电网技术*, 2009, 33(8): 1-7)
- [38] Stojmenovic I. Fog computing: a cloud to the ground support for smart things and machine-to-machine networks//*Proceedings of the Australasian telecommunication networks and applications conference*. Melbourne, Australia, 2014: 117-122
- [39] Bastug E, Bennis M, Médard M, Debbah M. Toward interconnected virtual reality: Opportunities, challenges, and enablers. *IEEE Communications Magazine*, 2017, 55(6): 110-117
- [40] Wu C, Zhang Y, Zhang L, Yang B, Chen X, Zhu W, Qiu L. ButterFly: Mobile collaborative rendering over GPU workload migration//*Proceedings of the international conference on computer communications*. Atlanta, USA, 2017: 1-9
- [41] Valavanis KP, Vachtsevanos GJ. *Handbook of unmanned aerial vehicles*. Dordrecht, Netherlands: Springer Publishing Company, 2015
- [42] Lin Y, Saripalli S. Sampling-based path planning for UAV collision avoidance. *IEEE Transactions on Intelligent Transportation Systems*, 2017, 18(11): 3179-3192
- [43] Stöcker C, Bennett R, Nex F, Gerke M, Zevenbergen J. Review of the Current State of UAV Regulations. *Remote Sensing*, 2017, 9(5): 1-26
- [44] Motlagh NH, Bagaia M, Taleb T. Uav-based iot platform: A crowd surveillance use case. *IEEE Communications Magazine*, 2017, 55(2): 128-34
- [45] Zeng Y, Zhang R, Lim TJ. Throughput maximization for UAV-enabled mobile relaying systems. *IEEE Transactions on Communications*, 2016, 64(12):4983-4996
- [46] Jeong S, Simeone O, Kang J. Mobile edge computing via a UAV-mounted cloudlet: Optimization of bit allocation and path planning. *IEEE Transactions on Vehicular Technology*, 2017, PP(99): 1-14
- [47] Mach P, Becvar Z. Mobile edge computing: a survey on architecture and computation offloading. *IEEE Communications Surveys and Tutorials*, 2017, 19(3): 1628-1656
- [48] ETSI. Mobile Edge Computing (MEC); Framework and Reference Architecture. ETSI GS MEC 003 V1.1.1, 2016
- [49] Dinh HT, Lee C, Niyato D, Wang P. A survey of mobile cloud computing: architecture, applications, and approaches. *Wireless Communications and Mobile Computing*, 2013, 13(18): 1587-1611
- [50] Dean J, Ghemawat S. MapReduce: simplified data processing on large clusters. *Communications of the ACM*, 2008, 51(1): 107-113
- [51] Hong K, Lillethun D, Ramachandran U, Ottenwalder B, Koldehofe B. Mobile fog: a programming model for large-scale applications on the Internet of Things//*Proceedings of the second ACM sigcomm workshop on mobile cloud computing*. Hong Kong, China, 2013: 15-20
- [52] Orsini G, Bade D, Lamersdorf W. Computing at the mobile edge: designing elastic android applications for computation offloading//*Proceedings of the IFIP wireless and mobile networking conference*. Munich, Germany, 2015: 112-119
- [53] Saurez E, Hong K, Lillethun D, Ramachandran U, Ottenw äder B. Incremental deployment and migration of geo-distributed situation awareness applications in the fog//*Proceedings of the 10th ACM international conference on distributed and event-based systems*. Irvine, USA, 2016: 258-269
- [54] Ottenw äder B, Koldehofe B, Rothermel K, Hong K, Lillethun D, Ramachandran U. Mcep: A mobility-aware complex event processing

- system. *ACM Transactions on Internet Technology (TOIT)*, 2014, 14(1): 1-24
- [55] Cheng B, Solmaz G, Cirillo F, Kovacs E, Terasawa K, Kitazawa A. FogFlow: Easy programming of IoT services over cloud and edges for smart cities. *IEEE Internet of Things Journal*, 2017, PP(99): 1-11
- [56] Willis D, Dasgupta A, Banerjee S. ParaDrop: a multi-tenant platform to dynamically install third party services on wireless gateways//*Proceedings of the 9th ACM workshop on mobility in the evolving internet architecture*. Maui, Hawaii, 2014: 43-48
- [57] Hu YC, Patel M, Sabella D, Sprecher N, Young V. Mobile edge computing—A key technology towards 5G ETSI White Paper, 2015
- [58] Ku I, Lu Y, Gerla M. Software-defined mobile cloud: architecture, services and use cases//*Proceedings of the wireless communications and mobile computing conference*. Nicosia, Cyprus, 2014: 1-6
- [59] Heller B, Sherwood R, McKeown N. The controller placement problem//*Proceedings of the first workshop on hot topics in software defined networks*. Helsinki, Finland, 2012: 7-12
- [60] Mijumbi R, Serrat J, Gorricho JL, Bouten N, De Turck F, Boutaba R. Network function virtualization: State-of-the-art and research challenges. *IEEE Communications Surveys & Tutorials*, 2016, 18(1): 236-62
- [61] Katsalis K, Nikaen N, Schiller E, Ksentini A, Braun T. Network slices toward 5G communications: Slicing the LTE network. *IEEE Communications Magazine*, 2017, 55(8): 146-54
- [62] Han B, Geopalakrishnan V, Ji LS, Lee S. Network function virtualization: challenges and opportunities for innovations. *IEEE Communications Magazine*, 2015, 53(2): 90-97
- [63] Satyanarayanan M, Chen Z, Ha K, Hu W, Richter W, Pillai P. Cloudlets: at the leading edge of mobile-cloud convergence//*Proceedings of the 6th international conference on mobile computing, applications and services*. Austin, USA, 2014: 1-9
- [64] Chen W, Xu L, Li G, Xiang Y. A lightweight virtualization solution for Android devices. *IEEE Transactions on Computers*, 2015, 64(10): 2741-2751
- [65] Fernandes E, Paupore J, Rahmati A, Simionato D, Conti M, Prakash A. FlowFence: Practical data protection for emerging IoT application frameworks//*Proceedings of the USENIX Security Symposium*. Austin, USA, 2016: 531-548
- [66] Liu W, Nishio T, Shinkuma R, Takahashi T. Adaptive resource discovery in mobile cloud computing. *Computer Communications*, 2014, 50: 119-129
- [67] Nishio T, Shinkuma R, Takahashi T, Mandayam BN. Service-oriented heterogeneous resource sharing for optimizing service latency in mobile cloud//*Proceedings of the first international workshop on mobile cloud computing and networking*. Bangalore, India, 2013: 19-26
- [68] Aazam M, Huh EN. Dynamic resource provisioning through Fog micro datacenter//*Proceedings of the international conference on pervasive computing and communication workshops*. Saint Louis, USA, 2015: 105-110
- [69] Skarlat O, Schulte S, Borkowski M, Leitner P. Resource provisioning for iot services in the fog//*Proceedings of the 9th international conference on service-oriented computing and applications*. Macau, China, 2016: 32-39
- [70] Zhan ZH, Liu XF, Gong YJ, Zhang J, Chung HS, Li Y. Cloud computing resource scheduling and a survey of its evolutionary approaches. *ACM Computing Surveys*, 2015, 47(4): 1-33
- [71] Kavvadia E, Sagiadinos S, Oikonomou K, Tsioutsoulis G, Aïssa S. Elastic virtual machine placement in cloud computing network environments. *Computer Networks*, 2015, 93: 435-447
- [72] Liu J, Zhang Y, Zhou Y, Zhang D, Liu H. Aggressive resource provisioning for ensuring QoS in virtualized environments. *IEEE Transactions on Cloud Computing*, 2015, 3(2): 119-131
- [73] Chun BG, Ihm S, Maniatis P, Naik M, Patti A. CloneCloud: elastic execution between mobile device and cloud//*Proceedings of the sixth conference on Computer systems*. Salzburg, Austria, 2011: 201-314
- [74] Barbera MV, Kosta S, Mei A, Stefa J. To offload or not to offload? the bandwidth and energy costs of mobile cloud computing//*Proceedings of the international conference on computer communications*. Turin, Italy, 2013: 1285-1293
- [75] Kosta S, Aucinas A, Hui P, Mortier R, Zhang XW. ThinkAir: dynamic resource allocation and parallel execution in the cloud for mobile code offloading//*Proceedings of the international conference on computer communications*. Orlando, USA, 2012: 945-953
- [76] Mac ís M, Guitart J. A genetic model for pricing in cloud computing markets//*Proceedings of the 2011 ACM symposium on applied computing*. Taiwan, China, 2011: 113-118
- [77] Xu H, Li B. Dynamic cloud pricing for revenue maximization. *IEEE Transactions on Cloud Computing*, 2013, 1(2): 158-171
- [78] Liu Y, Wang R, Han Z. Interference-constrained pricing for D2D networks. *IEEE Transactions on Wireless Communications*, 2017, 16(1): 475-486
- [79] Chen X, Proulx B, Gong X, Zhang J. Exploiting social ties for cooperative D2D communications: A mobile social networking case. *IEEE/ACM Transactions on Networking*, 2015, 23(5): 1471-1484
- [80] Jin Y, Liu F, Yi X, Chen M. Reducing cellular signaling traffic for heartbeat messages via energy-efficient D2D forwarding//*Proceedings of the 37th international conference on distributed computing systems*. Atlanta, USA, 2017: 1301-1311
- [81] Toosi AN, Calheiros RN, Buyya R. Interconnected cloud computing environments: challenges, taxonomy, and survey. *ACM Computing Surveys*, 2014, 47(1): 1-47
- [82] Ibrahim MH. Octopus: An edge-fog mutual authentication scheme. *International Journal of Network Security*, 2016, 18(6): 1089-1101
- [83] Bouzeffrane S, Mostefa AFB, Houacine F, Cagnon H. Cloudlets authentication in nfc-based mobile computing//*Proceedings of the 2nd IEEE international conference on mobile cloud computing, Service, and Engineering*. Oxford, UK, 2014: 267-272
- [84] Yang X, Huang X, Liu JK. Efficient handover authentication with user

- anonymity and untraceability for mobile cloud computing. *Future Generation Computer Systems*, 2016, 62: 190-195
- [85] Yu S, Wang C, Ren K, Lou W. Achieving secure scalable and fine-grained data access control in cloud computing//*Proceedings of the international conference on computer communications*. San Diego, USA, 2010: 1-9
- [86] Almutairi A, Sarfraz M, Basalamah S, Aref W, Ghafoor A. A distributed access control architecture for cloud computing. *IEEE Software*, 2012, 29(2): 36-44
- [87] Dsouza C, Ahn GJ, Taguinod M. Policy-driven security management for fog computing//*Proceedings of the 15th international conference on information reuse and integration*. San Francisco, USA, 2014: 16-23
- [88] Lu R, Liang X, Li X, Lin X, Shen X. Eppa: An efficient and privacy-preserving aggregation scheme for secure smart grid communications. *IEEE Transactions on Parallel and Distributed Systems*, 2012, 23(9): 1621-31
- [89] Dwork C. Differential privacy: a survey of results//*Proceedings of the international conference on theory and applications of Models of Computation*. Xi'an, China, 2008: 1-19
- [90] Yang R, Xu Q, Au MH, Yu Z, Wang H, Zhou L. Position based cryptography with location privacy: A step for Fog Computing. *Future Generation Computer Systems*, 2018, 78(2): 799-806
- [91] Wang Y, Xu D, He X, Zhang C, Li F, Xu B. L2P2: Location-aware location privacy protection for location-based services//*Proceedings of the international conference on computer communications*. Orlando, USA, 2012: 1996-2004
- [92] Yi S, Qin Z, Li Q. Security and privacy issues of fog computing: A survey//*Proceedings of the international conference on wireless algorithms, systems, and applications*. Guilin, China, 2015: 685-695
- [93] NowSecure. Mobile security report. Oak Park, Illinois, USA: NowSecure, 2016
- [94] Alrawais A, Alhothaily A, Hu C, Cheng X. Fog computing for the Internet of Things: Security and privacy issues. *IEEE Internet Computing*, 2017, 21(2): 34-42



ZHOU Yue-Zhi, born in 1975, Ph.D., professor. His research interests include ubiquitous computing, distributed system, mobile device and systems.

ZHANG Di, born in 1987, Ph.D. His research interests include cloud computing, distributed system, and mobile network.

Background

Cloud computing is a form of Internet-based computing paradigm that provides ubiquitous, on-demand access to a shared pool of configurable computing resources (e.g., computation, storage, application, and services). Since its introduction in 2007, cloud computing has attracted a lot of attention and has developed rapidly over the last decade. However, the centralized cloud computing is difficult to meet the requirements of some emerging applications and services, such as the Internet of Things, smart city, and smart grid. Therefore, in recent years, a variety of network computing paradigms, such as fog computing, multi-access edge computing, and edge computing are proposed by industry and academia. Although with different terms, they have the same basic idea: further expanding the cloud computing and moving the computing infrastructure from remote data centers to edge routers, based station and servers that are close to end-devices, thus overcoming the bottleneck of cloud computing and providing better performance and user experience. These post-cloud computing paradigms can be collectively referred to

as near-end cloud computing. Summarizing these near-end cloud computing paradigms makes it easier to understand the underlying causes of these paradigms, and to understand their application scenarios and the challenges they face in order to study them in depth.

This paper systematically summarizes the emerged network computing paradigms in recent years, and then tries to capture the basic features of these new paradigms. First, the main technology and application bottlenecks encountered in cloud computing are analyzed, which include device and network heterogeneity, end and near-end equipment idleness, and the complexity of application scenarios. Second, the conception and architecture of the recently proposed near-end cloud computing are analyzed and presented. Third, the analysis of three typical near-end cloud computing is given, i.e., fog computing, multi-access edge computing and edge computing. Then, the opportunities and challenges of near-end cloud computing are discussed. Finally, the future development of near-end cloud computing is prospected and the paper is

concluded.

This work is supported by the Tsinghua University Initiative Scientific Research Program (No. 20161080066).

《计算机学报》