



无线边缘智能助力智慧社会

王旨, 张兴, 王文博
(北京邮电大学, 北京 100876)

摘要: 为了满足增强移动带宽、海量机器类通信和超可靠低时延通信三大 5G 应用场景的技术需求, 无线边缘智能将计算、存储以及网络等核心能力下沉到接入网边缘, 并在网络边缘引入深度学习、人工智能等技术, 从而赋能数字化生活、数字化社会、数字化工业。首先, 阐明了无线边缘智能的概念、面临的挑战和机遇; 其次, 分析了无线边缘智能的平台部署; 最后, 探讨了无线边缘智能未来可能的发展方向。

关键词: 5G; 边缘智能; 无线网络

中图分类号: TN929.5

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2019044

Wireless edge intelligence empowers smart society

WANG Zhi, ZHANG Xing, WANG Wenbo
Beijing University of Posts and Telecommunications, Beijing 100876, China

Abstract: To satisfy the technical requirements of eMBB, mMTC and uRLLC in 5G application scenarios, wireless edge intelligence sinks core capabilities of computing, caching and network to the edge of radio access network, as well as introduces technologies such as deep learning/artificial intelligence into the edge cloud, thereby enabling digital life, digital society and digital industry. Firstly, the concept, challenges and opportunities of wireless edge intelligence were clarified. Moreover, the platform deployment of wireless edge intelligence was analyzed. Finally, the possible future deployment directions of wireless edge intelligence were discussed.

Key words: 5G, edge intelligence, wireless network

1 引言

纵览全球, 4G 服务已走进人们生活中的点点滴滴, 5G 时代即将到来, 回顾整个通信行业的发展, 伴随着层出不穷的新兴通信技术, 移动通信系统可谓面临着一次又一次的变革。从模拟化的

1G 和数字化的 2G, 到 IP 化的 3G, 图片、视频等高带宽多媒体通信快速发展, 速率逐步提升, 服务逐渐多样; 再到 CT 和 IT 融合的 4G, 实现了移动互联网业务全覆盖, 网速可达 3G 的 50 倍, 容量、稳定性、传输速率等方面呈跳跃式增长; 乃至现在蓄势待发的 5G, 已不局限于人与人互

收稿日期: 2019-01-07; 修回日期: 2019-02-24

基金项目: 国家自然科学基金资助项目 (No.61571054)

Foundation Item: The National Natural Science Foundation of China (No.61571054)

联,有望实现任何时间、任何地点、与任何人或物互联,容量更是提升到4G的1 000倍,速率提升10~100倍,端到端时延缩短至毫秒级,是真正意义上的万物深度融合网络^[1-4]。

IDC统计报告显示^[5],到2020年,将有超过500亿台终端与设备联网,海量物联产生的数据空前浩大,并且在很多场景下需要及时响应并快速做出决策,显然通过集中式云计算是不合理和难以实现的。在这些情况下,数据需要在靠近数据源的边缘侧进行处理,以提供敏捷联接、实时业务、数据优化、安全和隐私保护等能力^[6-8]。为了应对上述全新的机遇与挑战,边缘计算作为一种全新的技术架构应运而生,其在靠近人、物或数据源头的边缘侧,融合了网络、计算、存储、应用核心能力的开放平台,就近提供边缘智能服务,从而有望满足物联网时代更为苛刻的需求^[9]。边缘计算改变了传统移动通信系统中网络与业务分离的状态,将边缘业务平台下沉到接入网边缘,更加靠近用户终端,为移动用户就近提供业务计算和数据缓存能力,实现了网络从接入管道化向信息化使能平台的关键跨越,如图1所示,有效地利用了边缘服务平台所提供的存储、计算、网

络资源,能够进一步满足3GPP定义的5G三大典型场景对高带宽、本地管理及低时延的需求^[10-11]。

毫无疑问,人工智能正在成为一项改变各行各业的通用技术^[12]。此前,绝大部分的人工智能都依赖云端实现^[13];现在,部署在终端侧的人工智能正在开辟更为广阔的市场。除了智能手机等个人终端之外,靠近设备或数据源头的网络边缘,物联网终端将会成为更大的战场^[14-15]。值得指出的是,依靠云端实现的人工智能存在一些弊端:首先,个人数据上传到云端不利于隐私保护;其次,云端人工智能需要较高的带宽支持,功耗较高;最后,云端人工智能会有较高的时延性,很多实时物联网场景无法应用。显然,要在靠近终端的网络边缘侧部署一定的智能,才能在物联网边缘构建起健硕的应用生态。

边缘智能是未来趋势,其优势显著,具体如下。

- 网络时延优势:靠近终端的位置确保时延敏感类应用的实时性需求。
- 网络分布优势:部署于边缘的人工智能让终端设备直接得到AI分析结果,既减少了网络流量,又分流了云端集中推理的计算压力。

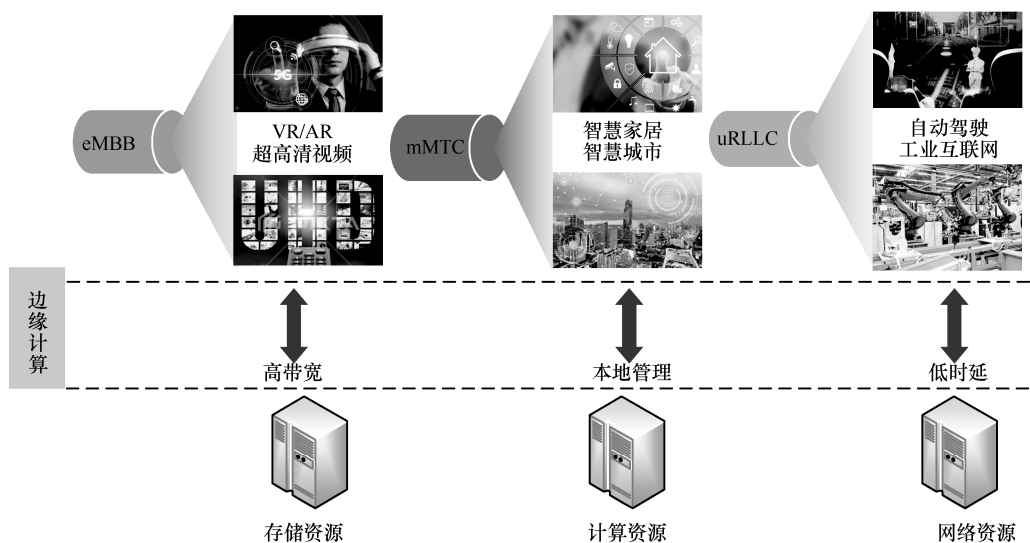


图1 边缘计算使能三大应用场景



- 安全性和可靠性提升：边缘智能不必将个人数据上传到云端，就地进行人工智能分析，有利于隐私保护，同时能提升对网络依赖方面的可靠性。
- 产业布局灵活：围绕边缘智能将会衍生出全新的生态，由于人工智能的各种算法发展日新月异、组合千变万化，边缘智能将会推动产业的进一步分工，产生算法提供商、终端生产商、边缘智能服务运营商等多重角色。

综上所述，5G 铺路，边缘计算使能人工智能驶入“快车道”，助力人工智能拥有更高的安全性、更低的功耗、更短的时延、更高的可靠性以及更低的带宽需求；同时，边缘计算的更高价值是边缘智能，是智能化云计算的落地部署，可使局域、边缘的物联网系统具备自治自律的行为能力和自给自足的算力。

本文首先介绍了边缘智能的基本概念，比较智能在无线网络中部署的 3 种方式，并梳理了边缘智能面临的种种挑战和机遇；然后针对实际搭建的边缘智能平台，分析其参考架构及核心功能组件；最后将边缘智能引入当下热门的关键技术

中，为后续边缘智能的落地提供理论参考。

2 浅谈无线边缘智能

2.1 基本概念

无线边缘智能指在无线网络边缘侧部署人工智能，即强调靠近数据源头，以减少智能化云计算服务交付的时延，实现真正意义上的无线状况智能感知、智能化快速决策和实时响应。

无线边缘智能逻辑架构如图 2 所示，逻辑上，边缘智能平台（edge intelligence platform, EIP）部署在传统接入网和核心网之间，使在移动网络边缘提供业务本地化和低时延内容分发服务成为可能^[16]。边缘智能平台通过透明解析由基站发给核心网的用户面数据分组，实现数据面报文的本地分流转发，从而近距离、低时延地为用户提供本地化服务，不仅缓解了用户续航性能不稳定的压力，而且为云端智能平台分担了计算和缓存任务；另一方面，控制面报文采用透传的方式传输给核心网，因此边缘智能平台对现有 LTE 网络无任何改造需求。

2.2 智能化在无线网络中部署的 3 种主要方式

根据智能化在无线网络中的部署位置，可分

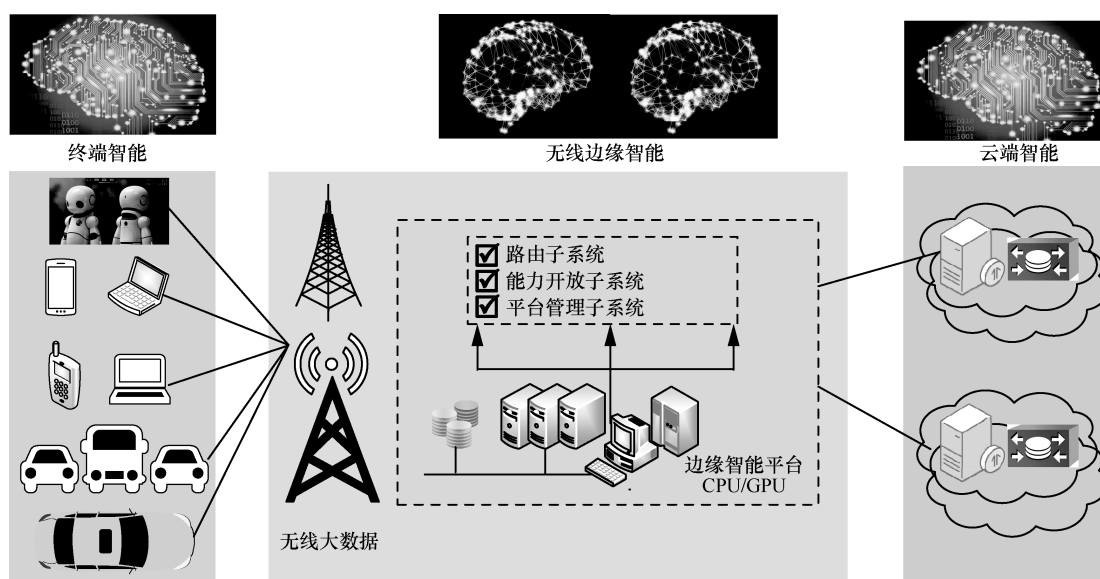


图2 无线边缘智能逻辑架构

表 1 3 种智能部署方式

	终端智能	云端智能	边缘智能
优点	针对设备级特定应用； 极低的数据获取时延； 部署灵活，可软可硬	丰富的计算与存储资源； 全局网络信息可汇总； 高度稳定的性能	融合信息的综合智能处理； 及时汇总用户、无线侧信息； 相较终端显著增强的运算与存储能力； 相较云计算明显降低的传输时延和响应时间
缺点	有限的计算与存储能力； 终端续航性能不稳定； 缺少网络的全局信息	较高的传输时延； 缺少用户情景信息； 易受回传链路限制	部署额外智能处理成本； 计算与存储依然受限

为终端智能、云端智能和边缘智能 3 种方式，见表 1，其中罗列了各种方式的优缺点。

显然，相较于终端智能，边缘智能的计算与缓存能力明显增强，可以将终端很难完成的计算复杂的任务卸载到边缘智能平台进行处理；而相较于云端智能，边缘智能优势明显，不仅可以感知无线侧信息、汇总用户相关信息，还能明显减少传输时延和响应时间，就近完成数据预处理及智能决策，为用户提供“贴地”智能云计算服务。然而值得指出，相比云端智能，边缘智能的计算与存储能力依然受限，且需要额外开销用于部署边缘侧的智能化平台。

2.3 边缘智能面临的挑战

随着 5G 大规模商用化的临近，边缘智能被运营商视为摆脱网络“被管道化”的有力支撑，并有望在 5G 时代成为对抗 OTT(over the top)的“杀手锏”。近两年，边缘智能产业生态逐渐构建，不过在形势大好的背后，边缘智能仍然面临着各种各样的难题。

其一，由于云计算服务的下沉，有些流量直接通过本地边缘智能平台流出。在传统核心网中，计算、控制、安全等均在核心网内部完成，如何对边缘智能平台的流量进行计费和控制，需要进行研究；同时边缘智能在实际的网络架构上有多种不同的部署规划，比如部署在无线接入云、边缘云及汇聚云，而不同的架构所面临的问题也略

有不同。其二，边缘智能所服务的对象以及场景较为多样化，如何使一套边缘智能平台适应多样化的第三方应用，也是目前面临的一个问题。其三，因为业务的碎片化，边缘智能可能针对单一场景或者两三个场景进行部署，不仅需要在整个部署能力特性上进行明确定义，且需要考虑如何进行敏捷智能化的运维。其四，如何更好地引入人工智能，发挥边缘计算的整体优势有待研究。其五，协调如何将边缘智能平台内部基于 OpenStack 的平台管理子系统进一步轻量化，并考虑将一些微服务架构（如 Docker、容器技术）迅速地引入边缘智能平台，从而降低管理开销。此外，边缘智能的商业运营模式、部署位置、自愈和自动扩缩容等问题仍需要得到进一步的明确和解决。

2.4 边缘智能的机遇

作为下一波计算浪潮，虽然边缘智能面临着很多挑战，但是从通信行业、互联网行业再到工业领域，普遍对边缘智能将要发挥的重大作用寄予了厚望。

首先，边缘智能缩短任务执行时延。移动应用任务处理时延包括传输时延、计算时延和通信时延。传统的移动云计算中，信息需要经过无线接入网、回传链路到达位于核心网的云服务器；而边缘智能将边缘服务器部署在无线接入网侧，缩短了计算服务器与移动设备之间的距离。由于传输距离的缩短，终端设备的任务卸载不需要经



过回传链路和核心网,减少了时延开销。另一方面,边缘服务器的计算处理能力远大于移动设备,从而大幅缩短了任务计算时延。显然,边缘智能传输距离短、协议扁平化的特点使其能满足 5G 网络的超低时延需求。

其次,边缘智能大幅提升网络能效。物联网设备可广泛应用到环境监测、人群感知、智能农业等多种场景,但部署的物联网设备大多无法通过电网供电,在设备电池能量有限的情况下,边缘智能缩短了边缘服务器与移动设备的距离,大幅节约了任务卸载、无线传输所损耗的能量,从而有效延长了物联网设备的使用周期。

此外,边缘智能提供更高的服务可靠性。边缘智能的服务器采用分布式部署,单个服务器服务规模小,不存储过多的有价值信息。因此,相较于移动云计算的数据中心,不易成为被攻击的目标,可以提供更可靠的服务。同时,多数移动边缘云服务器属于私有云,信息泄露风险低,使其具有更高的安全性。

3 无线边缘智能平台部署

无线边缘智能平台包含了三大基本功能组件:分别为路由子系统、网络能力开放子系统及平台管理子系统^[17]。其中,路由子系统在无线接入网、边缘智能平台及核心网之间提供网络连接和数据转发通路;网络能力开放子系统通过平台中间件提供的 API 向授权第三方开放无线网络能力(包括小区负载、无线信道质量、信干噪比、吞吐量及用户位置相关的信息等),有助于第三方灵活快速部署新兴的应用和接入网侧的服务;平台管理子系统负责调度和编排本地 IT 基础设施(即计算、存储和网络资源池),同时支撑应用认证和服务授权的工作。

具体来说,边缘智能参考架构包含以下核心模块:

- 数据分组封装/解封装模块和 DPDK^[18]快速

转发模块实现路由子系统中的数据面报文处理;

- 分流规则控制模块通过预先设置分流规则,将匹配分流规则的 GTP-U 数据分组转发至业务虚拟机;
- 域名解析模块存放查询过的域名及从何处获得域名映射信息的记录;
- 服务注册与生命周期管理模块实现边缘服务注册、发现、注销、信息更新等功能;
- RNIS 模块定时从定制化的开源基站获取需要的 RAN 侧信息,并保存在数据库中。

显然,数据分组封装/解封装模块、DPDK 快速转发模块、分流规则控制模块及域名解析模块包含在路由子系统中。因此,作为边缘智能平台部署中不可或缺的重要组成部分,路由子系统通过预先设置相应的分流规则(IP 地址五元组),对外提供网络流量控制并实现业务流量的本地导出。换句话说,路由子系统控制着无线接入网和核心网之间数据分组的走向,是实现智能化云计算下沉到接入网边缘的核心功能。

接下来,简要叙述基于无线边缘智能的蜂窝网络中本地数据分流方案,如图 3 所示。

步骤 1 当基站将用户请求发送至核心网时,边缘智能平台内部的路由子系统获得用户请求。

步骤 2 路由子系统中的数据分组封装/解封装模块将 GTP 分组解分组,以便查看目的 IP 地址是否与分流规则控制模块预先设置的分流规则相匹配。

步骤 3 若与分流规则匹配,用户请求重定向到业务虚拟机进行本地内容分发;若不匹配,用户数据分组被封装起来并传输到核心网。

值得指出的是,上述用户请求均为用户面数据,如图 3 实线箭头所示;而控制面数据采用透传的方式发送至核心网,如图 3 虚线箭头所示。显然,终端用户的控制依然由核心网负责,边缘智能平台保证了对终端和现有 LTE 网络是透明

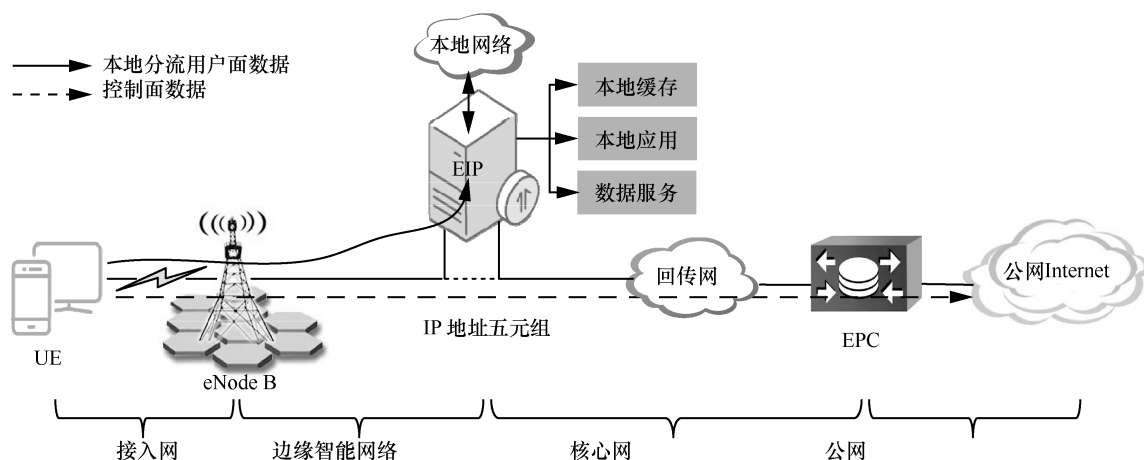


图3 边缘智能使能的LTE本地数据分流方案

的，即无需对现有终端及网络进行改造，从而使边缘智能方案更容易落地部署。

4 无线边缘智能的应用发展方向

无线边缘智能是使能网络大数据和智能应用的有效方式，因此，需要不断拓展当前对边缘智能的定义与应用范围，将边缘智能与5G时代众多时延敏感型去中心化的应用场景和关键技术相结合，有助于把边缘智能的优势发扬光大，进而助力智慧社会。

4.1 车联网

5G网络的高速率、低时延以及高可靠性将促进网联汽车的路况信息共享、无人高速驾驶以及车队编队等更多方面的高速发展，而云端智能和边缘智能的协作能够给网联车辆提供实时路况、前方道路信息以及周边行人信息等一系列重要的交通信息。精心设计的自动驾驶汽车能够在驾驶过程中通过将实时感知的数据卸载到边缘智能平台进行即时处理^[19]，同时又将数据发送到云端智能平台进行更多面向概述的工作，云端智能平台处理一段时间后最终通过分析数据向自动驾驶汽车发送一些新信息。因此，得益于大量的传感器数据以及边缘智能平台关键的本地即时处理能力^[20-21]，再加上连接到云端智能平台上大规模先进的数据分析工具，自动驾驶汽车迈进了“入5G网而知天下”

的时代。

4.2 区块链

区块链技术是一种互联网分布式数据库技术，具有去中心化、几乎不可篡改、公开透明等特性，每个节点的操作都记录在区块链上。区块链有种种好处，但缺点也异常明显，即需要具备很高的计算能力和存储容量，因此区块链在无线移动网络中仍未广泛应用^[22-23]。然而，鉴于边缘智能平台在接入网边缘融合计算、缓存和网络三大核心能力，相比终端设备具有显著增强的运算和存储能力，可以考虑把计算密集的挖矿任务卸载到邻近的边缘智能节点，从而解决了终端计算和存储能力受限导致“挖矿挖不动”的问题^[24]。因此，结合区块链和边缘智能的各自优势，可以在智慧农业系统、共享单车系统等时延敏感且需要安全监管的应用场景中大展拳脚。

4.3 智慧城市

随着传统行业自身数字化转型的需要，以及中国智慧城市、新型城镇化建设等国家战略的推动，积极地助燃了物联网产业的发展，未来的城市基础设施将进一步物联网化，各类传感器设备将无处不在，涉及交通出行、气象环保、工业制造、健康养老等诸多方面。为了应对物联网时代智慧城市场景下，数据之大量、很多场景需要及时响应、安全和隐私等问题的挑战，应该遵循“业



务应用在边缘智能平台,管理在云端智能”的模式^[25]。边缘智能端的业务规则是在云端通过机器学习的方式进行训练,训练好后推送到边缘智能端,本地执行。显然,边缘智能平台将承载更多的业务逻辑执行能力^[26],尤其针对那些需要及时响应的场景,而云端则需要 AI 来参与管理和协同。

4.4 工业互联网

工业互联网作为物联网在工业领域的应用,近年来得到了迅猛发展^[27]。在 2017 年的全国人民代表大会和中国人民政治协商会议上,“发展工业互联网平台”首次被写入政府工作报告中。工业互联网能够将人、数据、智能资产与设备相连接,在人工智能与大数据分析等技术的应用下,有效提升生产力与工作效率^[28],使能工业革命发展再上新台阶。相比传统泛在物联网设备,工业设备对计算的实时性、可靠性和安全性均提出了更高的要求。工业互联网中的边缘智能应用场景异常复杂^[29-30],考虑到边缘智能平台并不只是负责数据收集转发的网关,而是能够运行云端训练后的智能分析算法,从而提供智能化运算能力,且能够产生可操作的决策反馈,用于反向控制设备端。综上,无线边缘智能技术有望提升工业互联网海量数据在本地处理的速度和效率,减轻云端的负荷,进一步帮助用户过滤处理机密数据,脱敏后再上传到云端。

5 结束语

边缘智能作为使能 5G 三大典型应用场景的计算范式,把 IT 服务环境和云计算能力扩展到了移动网络边缘,必将与云端智能协同助力各行各业的数字化转型。本文从无线边缘智能的基本概念和与终端智能、云端智能的区别入手,介绍了边缘智能面临的挑战和机遇;详细描述了所提无线边缘智能平台的参考架构和核心功能;最后探讨了无线边缘智能未来可能的发展方向,为进一步推进边缘智能的部署和成熟提供了理论依托。

参考文献:

- [1] SHAFI M, MOLISCH A F, SMITH P J, et al. 5G: a tutorial overview of standards, trials, challenges, deployment and practice[J]. IEEE Journal on Selected Areas in Communications, 2017, 35(6): 1201-1221.
- [2] GUPTA A, JHA R K. A survey of 5G network: architecture and emerging technologies[J]. IEEE Access, 2015(3): 1206-1232.
- [3] PALATTELLA M R, DOHLER M, GRIECO A, et al. Internet of things in the 5G era: enablers, architecture, and business models[J]. IEEE Journal on Selected Areas in Communications, 2016, 34(3): 510-527.
- [4] AKPAKWU G A, SILVA B J, HANCKE G P, et al. A survey on 5G networks for the internet of things: communication technologies and challenges[J]. IEEE Access, 2018(6): 3619-3647.
- [5] IDC. Mobile edge computing-a gateway to 5G era[R]. 2016.
- [6] ABBAS N, ZHANG Y, TAHERKORDI A, et al. Mobile edge computing: a survey[J]. IEEE Internet of Things Journal, 2018, 5(1): 450-465.
- [7] MAO Y, YOU C, ZHANG J, et al. A survey on mobile edge computing: the communication perspective[J]. IEEE Communications Surveys & Tutorials, 2017, 19(4): 2322-2358.
- [8] MACH P, BECVAR Z. Mobile edge computing: a survey on architecture and computation offloading[J]. IEEE Communications Surveys & Tutorials, 2017, 19(3): 1628-1656.
- [9] ETSI. Mobile edge computing: a key technology towards 5G[S]. 2015.
- [10] ETSI. Developing software for multi-access edge computing[S]. 2017.
- [11] ETSI. MEC in 5G networks[S]. 2018.
- [12] GAO X, OUYANG D, SUN J, et al. AI in China: a survey[J]. IEEE Intelligent Systems, 2008, 23(6): 26-32.
- [13] LINTHICUM D S. Making sense of AI in public clouds[J]. IEEE Cloud Computing, 2017, 4(6): 70-72.
- [14] SHARIATMADARI H, RATASUK R, IRAJI S, et al. Machine-type communications: current status and future perspectives toward 5G systems[J]. IEEE Communications Magazine, 2015, 53(9): 10-17.
- [15] CHIANG M, ZHANG T. Fog and IoT: an overview of research opportunities[J]. IEEE Internet of Things Journal, 2016, 3(6): 854-864.
- [16] ETSI. MEC deployments in 4G and evolution towards 5G[S]. 2018.
- [17] YU Y. Mobile edge computing towards 5G: vision, recent progress, and open challenges[J]. China Communications, 2016, 13(Supplement2): 89-99.
- [18] BEGIN T, BAYNAT B, GALLARDO G A, et al. An accurate and efficient modeling framework for the performance evaluation of DPDK-based virtual switches[J]. IEEE Transactions on

- Network and Service Management, 2018, 10.1109/TNSM.2018.2874476.
- [19] LUO G, YUAN Q, ZHOU H, et al. Cooperative vehicular content distribution in edge computing assisted 5G-VANET[J]. China Communications, 2018, 15(7): 1-17.
- [20] HUANG X, YU R, KANG J, et al. Distributed reputation management for secure and efficient vehicular edge computing and networks[J]. IEEE Access, 2017(5): 25408-25420.
- [21] GARG S, SINGH A, BATRA S, et al. UAV-empowered edge computing environment for cyber-threat detection in smart vehicles[J]. IEEE Network, 2018, 32(3): 42-51.
- [22] XIONG Z, ZHANG Y, NIYATO D, et al. When mobile blockchain meets edge computing[J]. IEEE Communications Magazine, 2018, 56(8): 33-39.
- [23] LIU M, YU R, TENG Y, et al. Computing offloading and content caching in wireless blockchain networks with mobile edge computing[J]. IEEE Transactions on Vehicular Technology, 2018, 10.1109/TVT.2018.2866365.
- [24] JIAO Y, WANG P, NIYATO D, et al. Social welfare maximization auction in edge computing resource allocation for mobile blockchain[C]//2018 IEEE International Conference on Communications, May 20-24, 2018, Kansas City, MO, USA. Piscataway: IEEE Press, 2018: 1-6.
- [25] HOU W, NING Z, GUO L. Green survivable collaborative edge computing in smart cities[J]. IEEE Transactions on Industrial Informatics, 2018, 14(4): 1594-1605.
- [26] LI M, SI P, ZHANG Y. Delay-tolerant data traffic to software-defined vehicular networks with mobile edge computing in smart city[J]. IEEE Transactions on Vehicular Technology, 2018, 67(10): 9073-9086.
- [27] XU H, YU W, GRIFFITH D, et al. A survey on industrial internet of things: a cyber-physical systems perspective[J]. IEEE Access, 2018(6): 78238-78259.
- [28] PORAMBAGE P, OKWUIBE J, LIYANAGE M, et al. Survey on multi-access edge computing for internet of things realization[J]. IEEE Communications Surveys & Tutorials, 2018, 20(4): 2961-2991.
- [29] UGWUANYI E E, GHOSH S, IQBAL M, et al. Reliable resource provisioning using bankers' deadlock avoidance algorithm in MEC for industrial IoT[J]. IEEE Access, 2018(6): 43327-43335.
- [30] SUN W, LIU J, YUE Y, et al. Double auction-based resource allocation for mobile edge computing in industrial internet of things[J]. IEEE Transactions on Industrial Informatics, 2018, 14(10): 4692-4701.

[作者简介]



王旨(1991-),男,北京邮电大学泛网无线通信教育部重点实验室博士生,主要研究方向为无线通信。



张兴(1980-),男,北京邮电大学教授、博士生导师,主要从事无线通信的研究和教学工作。



王文博(1965-),男,北京邮电大学副校长、教授、博士生导师,长期从事无线系统的研究和教学工作。