



信息通信领域产学研合作特色期刊 | 十佳皖刊  
第三届国家期刊奖百种重点期刊 | 中国科技核心期刊

ISSN 1009-6868  
CN 34-1228/TN

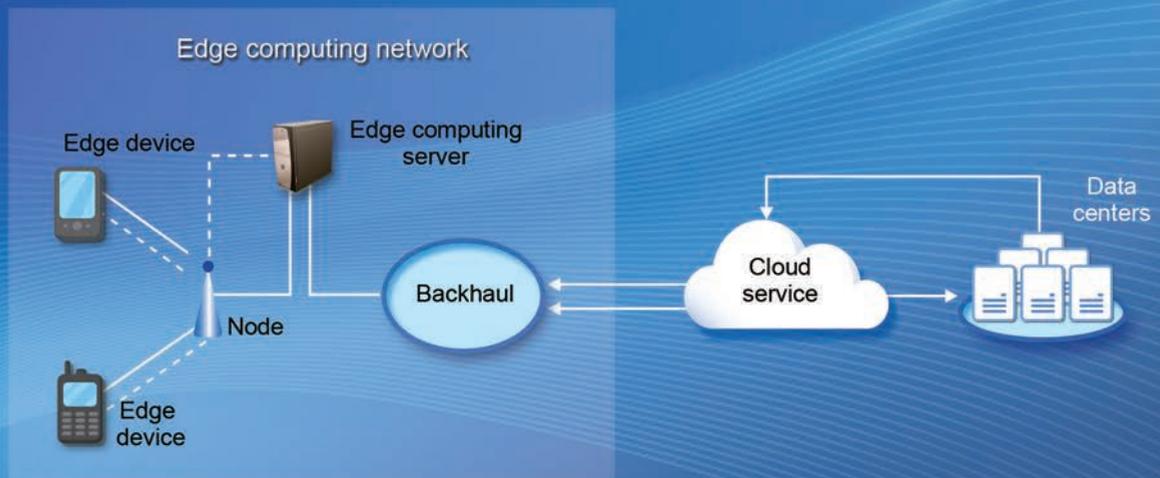
# 中兴通讯技术

## ZTE TECHNOLOGY JOURNAL

<http://tech.zte.com.cn>

2019年6月 • 第3期

### 专题：边缘计算技术及其应用



# 《中兴通讯技术》第8届编辑委员会成员名单

**顾问** 侯为贵（中兴通讯股份有限公司创始人） | 钟义信（北京邮电大学教授） | 陈锡生（南京邮电大学教授）

**主任** 陆建华（中国科学院院士）

**副主任** 徐子阳（中兴通讯股份有限公司总裁） | 糜正琨（南京邮电大学教授）

## 编委（按姓名拼音排序）

- |                           |                                |
|---------------------------|--------------------------------|
| <b>陈建平</b> 上海交通大学教授       | <b>谈振辉</b> 北京交通大学教授            |
| <b>陈前斌</b> 重庆邮电大学副校长      | <b>唐雄燕</b> 中国联通网络技术研究院首席科学家    |
| <b>葛建华</b> 西安电子科技大学教授     | <b>王文博</b> 北京邮电大学副校长           |
| <b>管海兵</b> 上海交通大学教授       | <b>王文东</b> 北京邮电大学教授            |
| <b>郭庆</b> 哈尔滨工业大学教授       | <b>王喜瑜</b> 中兴通讯股份有限公司执行副总裁     |
| <b>洪波</b> 中兴发展股份有限公司总裁    | <b>王翔</b> 中兴通讯股份有限公司高级副总裁      |
| <b>洪伟</b> 东南大学教授          | <b>卫国</b> 中国科学技术大学教授           |
| <b>纪越峰</b> 北京邮电大学教授       | <b>吴春明</b> 浙江大学教授              |
| <b>蒋林涛</b> 中国信息通信研究院科技委主任 | <b>邬贺铨</b> 中国工程院院士             |
| <b>李尔平</b> 浙江大学教授         | <b>徐安士</b> 北京大学教授              |
| <b>李红滨</b> 北京大学教授         | <b>徐子阳</b> 中兴通讯股份有限公司总裁        |
| <b>李建东</b> 合肥工业大学副校长      | <b>续合元</b> 中国信息通信研究院副总工        |
| <b>李军</b> 清华大学教授          | <b>薛一波</b> 清华大学教授              |
| <b>李乐民</b> 中国工程院院士        | <b>杨义先</b> 北京邮电大学教授            |
| <b>李融林</b> 华南理工大学教授       | <b>杨震</b> 南京邮电大学校长             |
| <b>李少谦</b> 电子科技大学教授       | <b>易芝玲</b> 中国移动研究院首席科学家        |
| <b>林晓东</b> 中兴通讯股份有限公司副总裁  | <b>张宏科</b> 北京交通大学教授            |
| <b>刘健</b> 中兴通讯股份有限公司高级副总裁 | <b>张平</b> 北京邮电大学教授             |
| <b>刘建伟</b> 北京航空航天大学教授     | <b>张云勇</b> 中国联通研究院院长           |
| <b>陆建华</b> 中国科学院院士        | <b>赵慧玲</b> 工业和信息化部科技委信息网络专家组组长 |
| <b>马建国</b> 广东工业大学教授       | <b>郑纬民</b> 清华大学教授              |
| <b>孟洛明</b> 北京邮电大学教授       | <b>钟章队</b> 北京交通大学教授            |
| <b>糜正琨</b> 南京邮电大学教授       | <b>周亮</b> 南京邮电大学教授             |
| <b>任品毅</b> 西安交通大学教授       | <b>朱近康</b> 中国科学技术大学教授          |
| <b>孙知信</b> 南京邮电大学教授       | <b>祝宁华</b> 中国科学院半导体研究所副所长      |

# 目次

中兴通讯技术 (ZHONGXING TONGXUN JISHU)  
总第146期 第25卷 第3期 2019年6月

## 专题:边缘计算技术及其应用

### 边缘计算综述:应用、现状及挑战 02

丁春涛,曹建农,杨磊,王尚广

### 边缘计算开源平台现状分析 08

梁家越,刘斌,刘芳

### 边缘存储的发展现状与挑战 15

刘铎,杨涓,谭玉娟

### 移动边缘计算:架构、应用和挑战 23

高志鹏,尧聪聪,肖楷乐

### 基于边缘计算的物端系统挑战与愿景 31

彭晓晖,徐志伟

### 万物互联背景下的边缘计算安全需求与挑战 37

马立川,裴庆祺,肖慧子

### 自然环境下基于异构多源的边缘计算公共 43

安全系统

孙辉,余莹

### 50 工业互联网智能制造边缘计算:现状与挑战

宋纯贺,曾鹏,于海斌

### 58 面向无人驾驶的边缘高精地图服务

唐洁,刘少山

### 68 边缘计算中数据驱动的智能应用:前景与挑战

秦永彬,韩蒙,杨清亮

## 专家论坛

### 77 面向5G的边缘计算及部署思考

马洪源

## 企业视界

### 82 B5G毫米波和太赫兹技术的背景、

应用和挑战

彭琳,段亚娟,别业楠

## 2019年第1—6期专题计划及策划人

### 1. 5G商用支撑理论及关键技术

中兴通讯股份有限公司执行副总裁 王喜瑜  
中兴通讯股份有限公司首席科学家 向际鹰

### 2. 云网一体化技术

中国联通网络技术研究院首席科学家 唐雄燕

### 3. 边缘计算技术及其应用

清华大学教授 郑纬民  
乔治亚州立大学教授 潘毅  
韦恩州立大学教授 施巍松

### 4. 5G通信安全技术

清华大学教授 李军

### 5. 新型光互连与光接入技术

北京大学教授 李红滨

### 6. 5G通信系统示范应用

中国信息通信研究院科技委主任 蒋林涛

# CONTENTS

ZTE TECHNOLOGY JOURNAL Vol. 25 No. 3 Jun. 2019

## Special Topic: Technologies and Applications of Edge Computing

- Edge Computing: Applications, State-of-the-Art and Challenges **02**  
DING Chuntao, CAO Jiannong, YANG Lei,  
WANG Shangguang
- The Present Situation of Open Source Platforms for Edge Computing **08**  
LIANG Jiayue, LIU Bin, LIU Fang
- A Survey on the Storage Issues in Edge Computing **15**  
LIU Duo, YANG Juan, TAN Yujuan
- Mobile Edge Computing: Architectures, Applications and Challenges **23**  
GAO Zhipeng, YAO Congcong, XIAO Kaile
- Challenges and Vision of Things System in the Edge Computing **31**  
PENG Xiaohui, XU Zhiwei
- Security Requirements and Challenges in Edge Computing for Internet of Everything **37**  
MA Lichuan, PEI Qingqi, XIAO Huizi
- Edge Computing-Based Systems for Multi-Source Heterogeneous in Public Safety **43**  
SUN Hui, YU Ying

- 50** Industrial Internet Intelligent Manufacturing Edge Computing: State-of-the-Art and Challenges  
SONG Chunhe, ZENG Peng, YU Haibin
- 58** Edge High-Precise Map Services for Autonomous Driving  
TANG Jie, LIU Shaoshan
- 68** Data-Driven Intelligent Application in Edge Computing: Prospects and Challenges  
QIN Yongbin, HAN Meng, YANG Qingliang

## Expert Forum

- 77** Thoughts on 5G Edge Computing and Deployment  
MA Hongyuan

## Enterprise View

- 82** Technical Background, Applications and Challenges of B5G Millimeter Wave and Terahertz Communication  
PENG Lin, DUAN Yajuan, BIE Yanan

期刊基本参数: CN 34-1228/TN\*1995\*b\*16\*86\*zh\*P\* ¥ 20.00\*15000\*12\*2019-06

## 敬告读者

本刊享有所发表文章的版权,包括英文版、电子版、网络版和优先数字出版版权,所支付的稿酬已经包含上述各版本的费用。未经本刊许可,不得以任何形式全文转载本刊内容;如部分引用本刊内容,须注明该内容出自本刊。

# 专题：边缘计算技术及其应用

## 专题策划人 郑伟民



清华大学计算机科学与技术系教授、博士生导师，曾任中国计算机学会理事长；长期从事计算机系统结构、大规模数据存储、高性能计算等领域的科研教学工作；主持并完成了“973”“863”、自然科学基金等科研项目36项，负责或参与工程项目11项；获国家科技进步一等奖1次、国家科技进步二等奖2次、国家技术发明奖1次；发表论文500余篇，出版著作10本。

## 专题策划人 潘毅



美国乔治亚州立大学计算机科学系州校董教授、系主任，曾经担任过中南大学长江学者讲座教授，清华大学IV访问讲席教授，北京大学、清华大学、浙江大学客座教授；主要研究领域包括分布式及云计算、网络技术和生物信息学；获2002国际计算机科学综合大会成就奖、2001国际情报科学财团(IISF)访问研究员资助、1999年代顿大学文理学院杰出研究奖等多个奖项；已发表SCI论文200余篇、国际会议论文150余篇，出版著作40余本。

## 专题策划人 施巍松



IEEE Fellow、ACM Distinguished Scientist、美国韦恩州立大学计算机科学系 Charles H. Gershenson 杰出教授，中国自动化学会边缘计算专业委员会副主任委员、中国边缘计算联盟专家委员会委员、ACM/IEEE 国际边缘计算大会(SEC)的创始人；边缘计算领域早期的提出者和倡导者之一，长期致力于边缘计算在工业界的推广；全国百篇优秀博士论文获得者、美国国家自然科学基金会杰出青年教授奖获得者等；所编著的《边缘计算》一书，是世界第一本关于边缘计算的教科书。

## 内容导读

近年来，以云计算模型为核心的集中式处理模式已无法高效处理边缘设备产生的数据，边缘计算因此应运而生。“边缘计算(edge computing)”一词由美国太平洋西北国家实验室的 R. LAMOTHE 于2013年在一个两页纸的内部报告中首次提出。2016年5月，美国韦恩州立大学施巍松教授团队给出了边缘计算的正式定义：“边缘计算是指在网络边缘执行计算的一种新型计算模型，边缘计算操作的对象包括来自于云服务的下行数据和来自于万物互联服务的上行数据，而边缘计算的边缘是指从数据源到云计算中心路径之间的任意计算和网络资源，是一个连续统。”边缘计算模型和云计算模型并不是取代的关系，而是相辅相成的关系。边缘计算可以有效减小计算系统的延迟，减少数据传输带宽，缓解云计算中心压力，提高可用性，并能够保护数据安全和隐私。得益于这些优势，边缘计算迅速发展，得到了学术界和工业界的重视。2016年，美国自

然科学基金委将边缘计算列为突出领域；同年10月，计算机协会(ACM)和电气和电子工程师协会(IEEE)开始联合举办国际边缘计算大会(SEC)；2017年，中国首届边缘计算研讨会在合肥召开；全球首部边缘计算专业书籍——《边缘计算》也于2018年1月出版。2016年11月，由中国科学院沈阳自动化研究所、中国信息通信研究院、英特尔等在北京成立了边缘计算产业联盟，以推进边缘计算产业的健康可持续发展。从云到边缘被 Gartner 认为是2018年十大战略性技术趋势之一。

本期专题中，我们邀请了从事边缘计算的教授、专家撰写了10篇文章，从多个维度对边缘计算进行探讨，希望能引起大家的兴趣和讨论。在此，对各位作者的大力支持和精心撰稿表示衷心的感谢！

郑伟民、潘毅、施巍松

2019年4月18日

# 边缘计算综述： 应用、现状及挑战

## Edge Computing: Applications, State-of-the-Art and Challenges



丁春涛/DING Chuntao<sup>1</sup>, 曹建农/CAO Jiannong<sup>2</sup>, 杨磊/YANG Lei<sup>3</sup>, 王尚广/WANG Shangguang<sup>1</sup>

(1. 北京邮电大学, 北京 100876; 2. 香港理工大学, 香港 999077; 3. 华南理工大学, 广东 广州 510006)

(1. Beijing University of Posts and Telecommunications, Beijing 100876, China;

2. Hong Kong Polytechnic University, Hongkong 999077, China; 3. South China University of Technology, Guangzhou 510006, China)

**摘要:**通过对边缘计算概念、典型应用场景、研究现状及关键技术等系统性的介绍,认为边缘计算的发展还处在初级阶段,在实际的应用中还存在很多问题需要解决研究,包括优化边缘计算性能、安全性、互操作性以及智能边缘操作管理服务。

**关键词:**边缘计算; 安全性; 互操作性

**Abstract:** Through the systematic introduction of the concept of edge computing, typical application scenarios, research status and key technologies, we believe that the development of edge computing is still in its early stage, and there are still many problems in practical applications that need to be solved, including optimization of edge computing performance, security, interoperability, and intelligent edge operations management services.

**Key words:** edge computing; security; interoperability

DOI: 10.12142/ZTETJ.201903001

网络出版地址: <http://kns.cnki.net/kcms/detail/34.1228.TN.20190605.1023.002.html>

网络出版日期: 2019-06-05

收稿日期: 2018-12-27

思科在2016—2021年的全球云指数中指出:接入互联网的设备数量将从2016年的171亿增加到271亿。每天产生的数据量也在激增,全球的设备产生的数据量从2016年的218 ZB增长到2021年的847 ZB。传统的云计算模型是将所有数据通过网络上传至云计算中心,利用云计算中心的超强计算能力来集中解决应用的计算需求问题。然而,云计算的集中处理模式在万物互联的背景下有3点不足。

(1)万物互联实时性需求。万物互联环境下,随着边缘设备数量

的增加,这些设备产生的数据量也在激增,导致网络带宽逐渐成为了云计算的一个瓶颈。例如文献[1]指出:波音787每秒产生的数据量超过5 GB,但飞机与卫星之间的带宽不足以支持实时数据传输。

(2)数据安全与隐私。随着智能家居的普及,许多家庭在屋内安装网络摄像头,直接将摄像头收集的视频数据上传至云计算中心会增加泄露用户隐私数据的风险。

(3)能耗较大。随着在云服务器运行的用户应用程序越来越多,未来大规模数据中心对能耗的需求

将难以满足。现有的关于云计算中心的能耗研究主要集中在如何提高能耗使用效率方面<sup>[2]</sup>。然而,仅提高能耗使用效率,仍不能解决数据中心巨大的能耗问题,这在万物互联环境下将更加突出。

针对于此,万物互联应用需求的发展催生了边缘计算模型。边缘计算模型是指在网络边缘执行计算的一种新型计算模型。边缘计算模型中边缘设备具有执行计算和数据分析的处理能力,将原有云计算模型执行的部分或全部计算任务迁移到网络边缘设备上,降低云服务器

的计算负载,减缓网络带宽的压力,提高万物互联时代数据的处理效率。边缘计算并不是为了取代云,而是对云的补充,为移动计算、物联网等相关技术提供一个更好的计算平台。

边缘计算模型成为新兴万物互联应用的支撑平台,目前已是大势所趋。本文中,我们从概念、关键技术、典型应用、现状趋势和挑战等几个方面对边缘计算的模型展开详细介绍,旨在为边缘计算研究者提供参考。

## 1 边缘计算的概念

对于边缘计算,不同的组织给出了不同的定义。美国韦恩州立大学计算机科学系的施巍松等人把边缘计算定义为:“边缘计算是指在网络边缘执行计算的一种新型计算模式,边缘计算中边缘的下行数据表示云服务,上行数据表示万物互联服务”<sup>[1]</sup>。边缘计算产业联盟把边缘计算定义为:“边缘计算是在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,就近提供边缘智能服务,满足行业数字在敏捷联接、实时业务、数据优化、应用智能、安全与隐私保护等方面的关键需求”。

因此,边缘计算是一种新型计算模式,通过在靠近物或数据源头的网络边缘侧,为应用提供融合计算、存储和网络等资源。同时,边缘计算也是一种使能技术,通过网络边缘侧提供这些资源,满足行业在敏捷联接、实时业务、数据优化、应用智能、安全与隐私保护等

方面的关键需求。

### 1.1 边缘计算的体系架构

边缘计算通过在终端设备和云之间引入边缘设备,将云服务扩展到网络边缘。边缘计算架构包括终端层、边缘层和云层。图1展示了边缘计算的体系架构。接下来我们简要介绍边缘计算体系架构中每层的组成和功能。

(1)终端层。终端层是最接近终端用户的层,它由各种物联网设备组成,例如传感器、智能手机、智能车辆、智能卡、读卡器等。为了延长终端设备提供服务的时间,则应该避免在终端设备上运行复杂的计算任务。因此,我们只将终端设备负责收集原始数据,并上传至上层进行计算和存储。终端层连接上一层主要通过蜂窝网络。

(2)边缘层。边缘层位于网络的边缘,由大量的边缘节点组成,通常包括路由器、网关、交换机、接入点、基站、特定边缘服务器等。这些边缘节点广泛分布在终端设备和云层之间,例如咖啡馆、购物中心、公交总站、街道、公园等。它们能够对

终端设备上传的数据进行计算和存储。由于这些边缘节点距离用户距离较近,则可以为运行对延迟比较敏感的应用,从而满足用户的实时性要求。边缘节点也可以对收集的数据进行预处理,再把预处理的数据上传至云端,从而减少核心网络的传输流量。边缘层连接上层主要通过因特网。

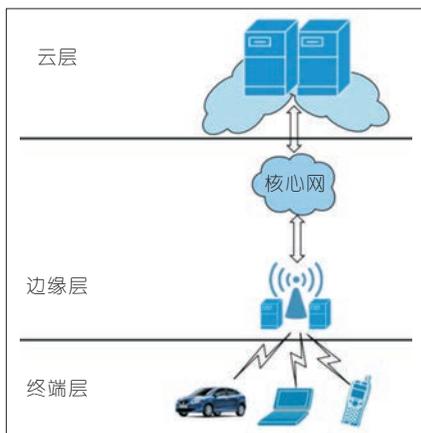
(3)云层。云层由多个高性能服务器和存储设备组成,它具有强大的计算和存储功能,可以执行复杂的计算任务。云模块通过控制策略可以有效地管理和调度边缘节点和云计算中心,为用户提供更好的服务。

### 1.2 边缘计算的范例

与边缘计算类似的范例,如雾计算、移动边缘计算等,虽然与边缘计算不尽相同,但它们在动机、节点设备、节点位置上与边缘计算范例类似。协同边缘计算<sup>[3]</sup>是一种新的计算范例,它使用边缘设备和路由器的网状网络来实现网络内的分布式决策。决策是在网络内部通过在边缘设备之间共享数据和计算而不是将所有数据发送到集中式服务器来完成的。这与通常执行集中计算的现有计算范例不同,并且诸如网关的边缘设备仅用于收集数据并将数据发送到服务器以进行处理。边缘计算与协同边缘计算的对比如表1所示。

### 1.3 边缘计算的优势

边缘计算模型将原有云计算中心的部分或全部计算任务迁移到数



▲图1 边缘计算体系结构

▼表1 边缘计算与协同边缘计算比较

|        | 边缘计算                   | 协同边缘计算           |
|--------|------------------------|------------------|
| 动机     | 支持物联网应用程序的移动性、位置感知和低延迟 | 允许多个服务提供商合作和共享数据 |
| 节点设备   | 路由器、交换机、网关             | 基站的服务器           |
| 节点位置   | 从终端设备到云                | 基站               |
| 软件架构   | 基于移动协调器                | 基于移动协调器          |
| 情境感知   | 中                      | 高                |
| 邻近跳数   | 一跳或多跳                  | 一跳               |
| 访问机制   | 蓝牙、Wi-Fi、移动网络          | 移动网络             |
| 节点之间通信 | 支持                     | 支持               |

据源附近,相比于传统的云计算模型,边缘计算模型具有实时数据处理和分析、安全性高、隐私保护、可扩展性强、位置感知以及低流量的优势。

(1)实时数据处理和分析。将原有云计算中心的计算任务部分或全部迁移到网络边缘,在边缘设备处理数据,而不是在外部数据中心或云端进行;因此提高了数据传输性能,保证了处理的实时性,同时也降低了云计算中心的计算负载。

(2)安全性高。传统的云计算模型是集中式的,这使得它容易受到分布式拒绝服务供给和断电的影响。边缘计算模型在边缘设备和云计算中心之间分配处理、存储和应用,使得其安全性提高。边缘计算模型同时也降低了发生单点故障的可能性。

(3)保护隐私数据,提升数据安全。边缘计算模型是在本地设备上处理更多数据而不是将其上传至云计算中心,因此边缘计算还可以减少实际存在风险的数据量。即使设备受到攻击,它也只会包含本地收集的数据,而不是受损的云计算

中心。

(4)可扩展性。边缘计算提供了更便宜的可扩展性路径,允许公司通过物联网设备和边缘数据中心的组合来扩展其计算能力。使用具有处理能力的物联网设备还可以降低扩展成本,因此添加的新设备都不会对网络产生大量带宽需求。

(5)位置感知。边缘分布式设备利用低级信令进行信息共享。边缘计算模型从本地接入网络内的边缘设备接收信息以发现设备的位置。例如导航,终端设备可以根据自己的实时位置把相关位置信息和数据交给边缘节点来进行处理,边缘节点基于现有的数据进行判断和决策。

(6)低流量。本地设备收集的数据可以进行本地计算分析,或者在本地设备上进行数据的预处理,不必把本地设备收集的所有数据上传至云计算中心,从而可以减少进入核心网的流量。

## 2 边缘计算的典型应用

边缘计算在很多应用场景下都取得了很好的效果。本节中,我们

将介绍基于边缘计算框架设计的几个新兴应用场景,部分场景在欧洲电信标准化协会(ETSI)白皮书中进行了讨论,如视频分析和移动大数据。还有一些综述论文<sup>[4,5]</sup>介绍了车辆互联、医疗保健、智能建筑控制、海洋监测以及无线传感器和执行器网络与边缘计算结合的场景。

(1)医疗保健。边缘计算可以辅助医疗保健,例如可以针对患有中风的患者辅助医疗保健。研究人员最近提出了一种名为U-fall的智能医疗基础设施,它通过采用边缘计算技术来利用智能手机<sup>[6]</sup>。在边缘计算的辅助下,U-fall借助智能设备传感器实时感应运动检测。边缘计算还可以帮助健康顾问协助他们的病人,而不受其地理位置的影响。边缘计算使智能手机能够从智能传感器收集患者的生理信息,并将其发送到云服务器以进行存储、数据同步以及共享。

(2)视频分析。在万物联网时代,用于监测控制的摄像机无处不在,传统的终端设备——云服务器架构可能无法传输来自数百万台终端设备的视频。在这种情况下,边缘计算可以辅助基于视频分析的应用。在边缘计算辅助下,大量的视频不用再全部上传至云服务器,而是在靠近终端设备的边缘服务器中进行数据分析,只把边缘服务器不能处理的小部分数据上传至云计算中心即可。

(3)车辆互联。通过互联网接入为车辆提供便利,使其能够与道路上的其他车辆连接。如果把车辆收集的数据全部上传至云端处理会

造成互联网负载过大,导致传输延迟,因此,需要边缘设备其本身具有处理视频、音频、信号等数据的能力。边缘计算可以为这一需要提供相应的架构、服务、支持能力,缩短端到端延迟,使数据更快地被处理,避免信号处理不及时而造成车祸等事故。一辆车可以与其他接近的车辆通信,并告知他们任何预期的风险或交通拥堵。

(4)移动大数据分析。无处不在的移动终端设备可以收集大量的数据,大数据对业务至关重要,因为它可以提取可能有益于不同业务部门的分析和有用信息。大数据分析是从原始数据中提取有意义的信息的过程。在移动设备附近实施部署边缘服务器可以通过网络高带宽和低延迟提升大数据分析。例如,首先在附近的边缘服务器中收集和析大数据,然后将大数据分析的结果传递到核心网络以进一步处理,从而减轻核心网络的压力。

(5)智能建筑控制。智能建筑控制系统由部署在建筑物不同部分的无线传感器组成。传感器负责监测和控制建筑环境,例如温度、气体水平或湿度。在智能建筑环境中,部署边缘计算环境的建筑可以通过传感器共享信息并对任何异常情况做出反应。这些传感器可以根据其他无线节点接收的集体信息来维持建筑气氛。

(6)海洋监测控制。科学家正在研究如何应对任何海洋灾难性事件,并提前了解气候变化。这可以帮助人们快速采取应对措施,从而减轻灾难性事件造成的严重后果。

部署在海洋中某些位置的传感器大量传输数据,这需要大量的计算资源和存储资源。而利用传统的云计算中心来处理接收到的大量数据可能会导致预测传输的延迟。在这种情况下,边缘计算可以发挥重要作用,通过在靠近数据源的地方就近处理,从而防止数据丢失或传感器数据传输延迟。

(7)智能家居。随着物联网技术的发展,智能家居系统得到进一步发展,其利用大量的物联网设备实时监测控制家庭内部状态,接收外部控制命令并最终完成对家居环境的调控,以提升家居安全性、便利性、舒适性。由于家庭数据的隐私性,用户并不总是愿意将数据上传至云端进行处理,尤其是一些家庭内部视频数据。而边缘计算可以将家庭数据处理推送至家庭内部网关,减少家庭数据的外流,降低数据外泄的可能性,提升系统的隐私性。

(8)智慧城市。预测显示:一个百万人口的城市每天将可能会产生200 PB的数据<sup>[7]</sup>。因此,应用边缘计算模型,将数据在网络边缘处理是一个很好的解决方案。例如,在城市路面检测中,在道路两侧路灯上安装传感器收集城市路面信息,检测空气质量、光照强度、噪声水平等环境数据,当路灯发生故障时能够即时反馈给维护人员,同时辅助健康急救和公共安全领域。

### 3 边缘计算现状和关键技术

目前,边缘计算的发展仍然处于初期阶段。随着越来越多的设备联网,边缘计算得到了来自工业界

和学术界的广泛重视和一致认可。本节中,我们主要从工业界和学术界的角度介绍边缘计算的现状。

#### 3.1 工业界

在工业界中,亚马逊、谷歌和微软等云巨头正在成为边缘计算领域的领先者。亚马逊的AWS Greengrass服务进军边缘计算领域,走在了行业的前面。AWS Greengrass将AWS扩展到设备上,这样本地生成的数据就可以在本地设备上处理。微软在这一领域也有大动作,该公司计划未来4年在物联网领域投入50亿美元,其中包括边缘计算项目。谷歌宣布了2款新产品,意在帮助改善边缘联网设备的开发。它们分别是硬件芯片Edge张量处理单元(TPU)和软件堆栈Cloud物联网(IoT)Edge。涉足边缘计算领域的并不只是这3大云巨头。2015年,思科、ARM、英特尔、微软、普林斯顿大学联合成立了开放雾计算(OpenFog)联盟;2016年11月30日,在北京正式成立了产学研结合的边缘计算产业合作平台,推动运行技术(OT)和信息与通信技术(ICT)产业开放协作,引领边缘计算产业蓬勃发展,深化行业数字化转型。

#### 3.2 学术界

学术界也展开了关于边缘计算的研究,边缘计算顶级年会电气和电子工程师协会(IEEE)/国际计算机协会(ACM)边缘计算研讨会(SEC)、IEEE国际分布式计算系统会议(ICDCS)、国际计算机通信会

议(INFOCOM)等重大国际会议都开始增加边缘计算的分会和专题研讨会。涉及主要关键技术及研究热点如下:

(1) 计算卸载。计算卸载是指终端设备将部分或全部计算任务卸载到资源丰富的边缘服务器,以解决终端设备在资源存储、计算性能以及能效等方面存在的不足。计算卸载的主要技术是卸载决策。卸载决策主要解决的是移动终端如何卸载计算任务、卸载多少以及卸载什么的问题。根据卸载决策的优化目标将计算卸载分为以降低时延为目标、以降低能量消耗为目标以及权衡能耗和时延为目标的3种类型。

(2) 移动性管理。边缘计算依靠资源在地理上广泛分布的特点来支持应用的移动性,一个边缘计算节点只服务周围的用户。云计算模式对应用移动性的支持则是服务器位置固定,数据通过网络传输到服务器,所以在边缘计算中应用的移动管理是一种新模式。主要涉及2个问题:资源发现,即用户在移动的过程中需要快速发现周围可以利用的资源,并选择最合适的资源。边缘计算的资源发现需要适应异构的资源环境,还需要保证资源发现的速度,才能使应用不间断的为用户提供服务。另一个问题是资源切换,即当用户移动时,移动应用使用的计算资源可能会在多个设备间切换。资源切换要将服务程序的运行现场迁移,保证服务连续性是边缘计算研究的一个重点。一些应用程序期望在用户位置改变之后继续为用户提供服务。边缘计算资源的异

构性与网络的多样性,需要迁移过程自适应设备计算能力与网络带宽的变化。文献[8]中,作者通过选择性地虚拟机迁移到最佳位置来优化迁移增益和迁移成本间的权衡。

除了以上2个关键技术,边缘计算研究热点还包括网络控制、内容缓存、内容自适应、数据聚合以及安全卸载等问题。在网络控制方面,文献[9]的作者提出了一种有效的工作负载切片方案,用户使用软件定义网络处理多边缘云环境中的数据密集型应用程序。在内容缓存方面,文献[10]的作者提出了一种用于自动驾驶服务的两级边缘计算框架,以便充分利用无线边缘的智能来协调内容传输。在内容适应方面,文献[11]的作者介绍了一种用于在多用户移动网络中优化基于HTTP的多媒体传送的新颖架构。在数据聚合方面,文献[12]的作者提出了混合整数规划公示和算法,用于物联网边缘网络中传感器测量数据的能量最优路由和多宿聚合问题,以及联合聚合和传播。在安全卸载方面,文献[13]的作者提出了一种名为MECPASS的新型协作DoS防御架构,以减轻来自移动设备的攻击流量。

#### 4 挑战

目前边缘计算已经得到了各行各业的广泛重视,并且在很多应用场景下开花结果;但边缘计算的实际应用还存在很多问题<sup>[9]</sup>需要研究。本文中,我们对其中的几个主要问题进行分析,包括优化边缘计算性能、安全性、互操作性以及智能

边缘操作管理服务。

(1) 优化边缘计算性能。在边缘计算架构中,不同层次的边缘服务器所拥有的计算能力有所不同,负载分配将成为一个重要问题。用户需求、延时、带宽、能耗及成本是决定负载分配策略的关键指标。针对不同工作负载,应设置指标的权重和优先级,以便系统选择最优分配策略。成本分析需要在运行过程中完成、分发负载之间的干扰和资源使用情况,都对边缘计算架构提出了挑战。

(2) 安全性。边缘计算的分布式架构增加了攻击向量的维度,边缘计算客户端越智能,越容易受到恶意软件感染和安全漏洞攻击。在边缘计算架构中,在数据源的附近进行计算是保护隐私和数据安全的一种较合适的方法。但由于网络边缘设备的资源有限,对于有限资源的边缘设备而言,现有数据安全的保护方法并不能完全适用于边缘计算架构。而且,网络边缘高度动态的环境也会使网络更加易受攻击和难以保护。

(3) 互操作性。边缘设备之间的互操作性是边缘计算架构能够大规模落地的关键。不同设备商之间需要通过制定相关的标准规范和通用的协作协议,实现异构边缘设备和系统之间的互操作性。

(4) 智能边缘操作管理服务。网络边缘设备的服务管理在物联网环境中需要满足识别服务优先级,灵活可扩展和复杂环境下的隔离线。在传感器数据和通信不可靠的情况下,系统如何通过利用多维参

考数据源和历史数据记录,提供可靠的服务是目前需要关注的问题。

## 5 结束语

本文主要从基本概念、应用场景、研究现状和关键技术、存在的挑战方面对边缘计算模型进行了系统性介绍。边缘计算的核心思想是为应用在网络边缘侧提供计算、存储和网络等资源,是一种新的计算架构。边缘计算架构可以满足用户对延迟敏感应用的需求和减少核心网络的负载压力。值得注意的是,单个边缘节点计算和存储资源有限且安全性低于云计算中心,如何实现边缘节点之间的安全、高性能协作和智能管理是目前亟待探索和研究的问题。

### 参考文献

- [1] 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. DOI:10.7554/issn1000-1239.2019.2018076
- [2] GAO Y Q, GUAN H B, QI Z W, et al. Service Level Agreement Based Energy-Efficient Resource Man agreement in Cloud Data Centers[J]. Computers & Electrical Engineering, 2014, 40(5): 1621-1633. DOI: 10.1016/j.compeleceng.2013.11.001
- [3] SAHNI Y, CAO J N, ZHANG S G, et al. Edge Mesh: A New Paradigm to Enable Distributed Intelligence in Internet of Things[J]. IEEE Access, 2017,(5):16441-16458. DOI: 10.1109/

- ACCESS.2017.2739804
- [4] ABBAS N, ZHANG Y, TAHERKORDI A, et al. Mobile Edge Computing: A Survey[J]. IEEE Internet of Things Journal, 2018, 5(1): 450-465. DOI:10.1109/jiot.2017.2750180
- [5] 施巍松,孙辉,曹杰,等. 边缘计算:万物互联时代新型计算模型[J]. 计算机研究与发展, 2017, 54(5): 907-924. DOI:10.7544/issn1000-1239.2017.20160941
- [6] CAO Y, CHEN S Q, HOU P, et al. FAST: A Fog Computing Assisted Distributed Analytics System to Monitor Fall for Stroke Mitigation [C]//2015 IEEE International Conference on Networking, Architecture and Storage (NAS). USA: IEEE, 2015: 2-11. DOI:10.1109/NAS.2015.7255196
- [7] Cisco Visual Networking. Cisco Global Cloud Index: Forecast and Methodology 2015-2020, CISCO White Paper[R].2015
- [8] SUN X, ANSARI N. PRIMAL: PRofit Maximization Avatar Placement for Mobile Edge Computing[C]//2016 IEEE International Conference on Communications (ICC). Malaysia: IEEE, 2016: 1-6. DOI:10.1109/ICC.2016.7511131
- [9] AUJLA G S, KUMAR N, ZOMAYA A Y, et al. Optimal Decision Making for Big Data Processing at Edge-Cloud Environment: An SDN Perspective [J]. IEEE Transactions on Industrial Informatics, 2018, 14(2): 778-789. DOI:10.1109/tii.2017.2738841
- [10] YUAN Q, ZHOU H B, LI J L, et al. Toward Efficient Content Delivery for Automated Driving Services: An Edge Computing Solution [J]. IEEE Network, 2018, 32(1): 80-86. DOI:10.1109/mnet.2018.1700105
- [11] FAJARDO J O, TABOADA I, LIBERAL F. Improving Content Delivery Efficiency through Multi-Layer Mobile Edge Adaptation [J]. IEEE Network, 2015, 29(6): 40-46. DOI: 10.1109/mnet.2015.7340423
- [12] FITZGERALD E, PIORO M, TOMASZEWSKI A. Energy-Optimal Data Aggregation and Dissemination for the Internet of Things[J]. IEEE Internet of Things Journal, 2018, 5(2): 955-969. DOI:10.1109/jiot.2018.2803792
- [13] NGUYEN V L, LIN P C, HWANG R H. MECPASS: Distributed Denial of Service Defense Architecture for Mobile Networks [J]. IEEE Network, 2018, 32(1): 118-124. DOI:10.1109/mnet.2018.1700140

### 作者简介



丁春涛,北京邮电大学网络技术研究院在读博士生;主要研究领域为边缘计算、云计算、机器学习。



曹建农,香港理工大学教授、博士生导师,IEEE Fellow,并担任中山大学、上海交通大学等多所重点大学的客座教授/兼职教授和博士生导师;主要研究领域为并行和分布式计算、无线网络和移动计算、大数据和云计算、普适计算、容错计算等;主持或参与开发超过90个科研项目,作为第一负责人承担并完成超过45个跨学科和全球合作科研项目。



杨磊,华南理工大学副教授;主要研究领域为移动云计算、边缘计算、物联网、大数据处理等;现主持国家自然科学基金、中央高校业务经费、CCF-腾讯犀牛鸟创意基金等项目;在国际主要期刊和会议上发表论文30余篇。



王尚广,北京邮电大学教授、博士生导师,担任国际期刊《International Journal of Web Science》的主编及2本SCI期刊的编委;主要研究领域为服务计算、边缘计算、云计算、群智计算等;主持完成国家/省部级项目4项,参与国家重点研发计划、国家“973”课题、“863”课题等6项;发表论文60余篇。

# 边缘计算开源平台 现状分析

## The Present Situation of Open Source Platforms for Edge Computing



梁家越/LIANG Jiayue<sup>1</sup>, 刘斌/LIU Bin<sup>2</sup>, 刘芳/LIU Fang<sup>1</sup>

(1. 中山大学, 广东广州 510006;

2. 中国科学院电子学研究所, 北京 100190)

(1. Sun Yat-Sen University, Guangzhou 510006, China;

2. Institute of Electronics, Chinese Academy of Sciences, Beijing 100190, China)

**摘要:** 给出了边缘计算平台的一般性功能框架,并基于平台的设计目标和部署方式,对当前典型的边缘计算开源平台进行了分类。从设计目标、目标用户、可扩展性、系统特点和应用领域等方面,对 EdgeX Foundry、Apache Edgent、Azure 物联网(IoT) Edge、CORD 和 Akraino Edge Stack 等典型开源平台进行了对比和分析。最后介绍了可用于构建边缘计算平台的一些开源软件,并讨论了边缘计算平台的未来发展方向和研究热点。

**关键词:** 边缘计算; 开源平台; 物联网; 开源软件

**Abstract:** In this paper, the general functional framework of edge computing platforms is analyzed, and the category of these platforms based on their design target and deployment characteristic is discussed. The existing typical open source edge computing platforms such as EdgeX Foundry, Apache Edgent, Azure Internet of Things (IoT) Edge, CORD and Akraino Edge Stack are introduced with comparison from the perspectives of design target, target user, scalability, system characteristic and application area. Finally, some open source software that can be used to build edge computing platform is introduced, and the future trends and research topics of edge computing platform are discussed.

**Key words:** edge computing; open source platform; IoT; open source software

DOI: 10.12142/ZTETJ.201903002

网络出版地址: <http://kns.cnki.net/kcms/detail/34.1228.TN.20190520.1814.002.html>

网络出版日期: 2019-05-20

收稿日期: 2018-12-30

随着万物互联的趋势不断加深,网络边缘产生的数据在急剧增长。受限于网络带宽增长速度远远跟不上数据的增长数据,传统云计算模式需要解决带宽和延迟这两大瓶颈,边缘计算<sup>[1]</sup>应运而生。边缘计算中的边缘(edge)指的是网络边缘上的计算和存储资源,这里的网络边缘与数据中心相对,无论是从地理距离还是网络距离上来看都更贴近用户<sup>[2]</sup>。作为一种新的计算范式,边缘计算将计算任务

部署于接近数据产生源的网络边缘,利用边缘资源为用户提供大量服务或功能接口,大大减少上传至云数据中心的数据量,缓解网络带宽压力,同时可以更好地解决数据安全 and 隐私问题。边缘计算平台对边缘计算领域的推广和发展有着重要的意义和影响<sup>[3]</sup>。

### 1 边缘计算开源平台概述

边缘计算系统是一个分布式系统范例,在具体实现过程中需要将

其落地到一个计算平台上,各个边缘平台之间如何相互协作提高效率,如何实现资源的最大利用率,对设计边缘计算平台、系统和接口带来挑战。例如,网络边缘的计算、存储和网络资源数量众多但在空间上分散,如何组织和统一管理这些资源,是一个需要解决的问题。在边缘计算的场景下,尤其是物联网,诸如传感器之类的数据源,其软件和硬件以及传输协议等具有多样性,如何方便有效地从数据源中采集数

据也是一个需要考虑的问题。此外,在网络边缘的计算资源并不丰富的条件下,如何高效地完成数据处理任务也是需要解决的问题。

目前,边缘计算平台的发展方兴未艾。由于针对的问题不同,各边缘计算平台的设计多种多样,但也不失一般性。边缘计算平台的一般性功能框架如图1所示。在该框架中,资源管理功能用于管理网络边缘的计算、网络和存储资源。设备接入和数据采集分别用于接入设备和从设备中获取数据。安全管理用于保障来自设备的数据的安全。平台管理功能用于管理设备和监测控制边缘计算应用的运行情况。

各边缘计算平台的差异可从以下方面进行对比和分析:

(1)设计目标。边缘计算平台的设计目标反映了其所针对解决的问题领域,并对平台的系统结构和功能设计有关键性的影响。

(2)目标用户。在现有的各种边缘计算平台中,有部分平台是提供给网络运营商以部署边缘云服

务;有的平台则没有限制,普通用户可以自行在边缘设备上部署使用。

(3)可扩展性。为满足用户应用动态增加和删除的需求,边缘计算平台需要具有良好的可扩展性。

(4)系统特点。面向不同应用领域的边缘计算开源平台具有不同的特点,而这些特点能为边缘计算应用的开发或部署带来方便。

(5)应用场景。常见的应用领域包括智能交通、智能工厂和智能家居等多种场景,还有增强现实(AR)/虚拟现实(VR)应用、边缘视频处理和无人车等对响应时延敏感的应用场景。

根据边缘计算平台的设计目标和部署方式,可将目前的边缘计算开源平台分为3类:面向物联网端的边缘计算开源平台、面向边缘云服务的边缘计算开源平台、面向云边融合的边缘计算开源平台。

## 2 面向物联网端的边缘计算开源平台

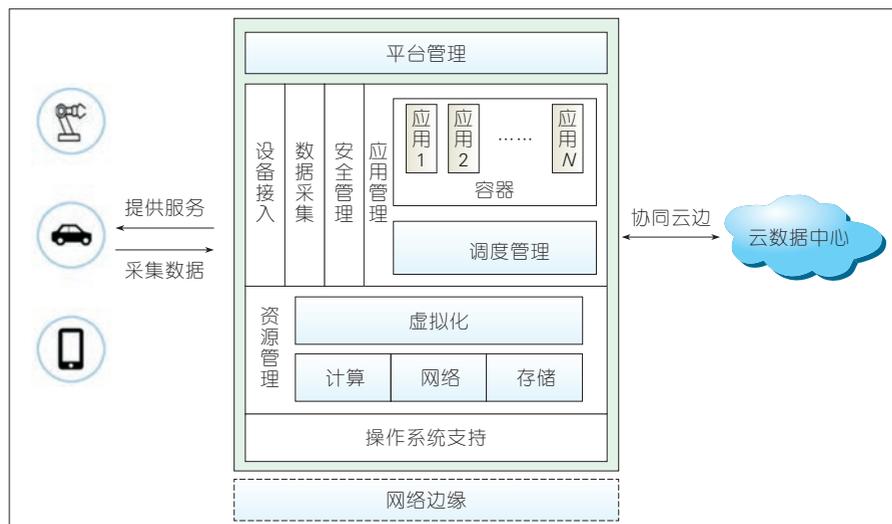
面向物联网端的边缘计算开源

平台,致力于解决在开发和部署物联网应用的过程中存在的问题,例如设备接入方式多样性问题等。这些平台部署于网关、路由器和交换机等边缘设备,为物联网边缘计算应用提供支持。代表性的2个平台是Linux基金会发布的EdgeX Foundry<sup>[4]</sup>和Apache软件基金会的Apache Edgent<sup>[5]</sup>。

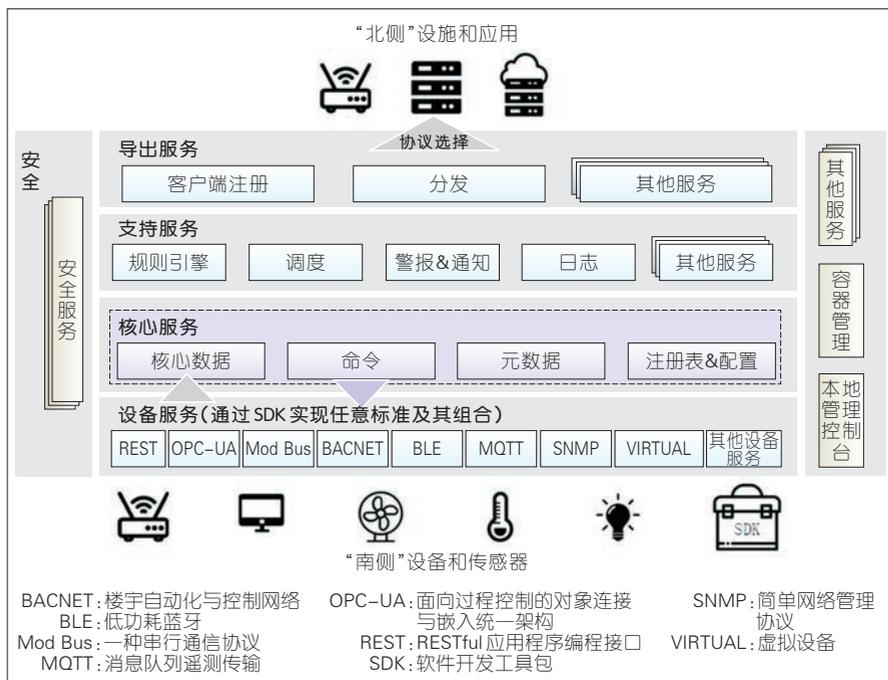
### 2.1 EdgeX Foundry

EdgeX Foundry是一个面向工业物联网边缘计算开发的标准化互操作性框架<sup>[4]</sup>,部署于路由器和交换机等边缘设备上,为各种传感器、设备或其他物联网器件提供即插即用功能并做管理,进而收集和分析它们的数据,或导出至边缘计算应用或云计算中心做处理。EdgeX Foundry针对的问题是物联网器件的互操作性。目前,具有大量设备的物联网产生大量数据,迫切需要结合边缘计算的应用;但物联网的软硬件和接入方式的多样性给数据接入功能带来困难,影响了边缘计算应用的部署。EdgeX Foundry的主旨是简化和标准化工业物联网边缘计算的架构,创建一个围绕互操作性组件的生态系统<sup>[4]</sup>。

图2展示了EdgeX Foundry的架构<sup>[4]</sup>。在图中的最下方是“南侧”,指的是所有物联网器件,以及与这些设备、传感器或其他物联网器件直接通信的边缘网络。在图的最上方是“北侧”,指的是云计算中心或企业系统,以及与云中心通信的网络部分。南侧是数据产生源,而北侧收集来自南侧的数据,并对



▲图1 边缘计算平台一般性功能框架



▲图2 EdgeX Foundry的架构图

数据进行存储、聚合和分析。如图2所示,EdgeX Foundry位于南侧和北侧两者之间,由一系列微服务组成,而这些微服务可以被分成4个服务层和2个底层增强系统服务。微服务之间通过一套通用的Restful应用程序编程接口(API)进行通信。

(1)设备服务层。设备服务层主要提供设备接入的功能,由多个设备服务组成。每个设备服务是用户根据设备服务软件开发工具包(SDK)编写生成的一个微服务。EdgeX Foundry使用设备文件去定义一个南侧设备的相关信息,包括源数据格式、存储在EdgeX Foundry中的数据格式,以及对该设备的操作命令等信息。设备服务将来自设备的数据进行格式转换,并发送至核心服务层。目前,EdgeX Foundry提供了消息队列遥测传输协议

(MQTT)、ModBus串行通信协议和低功耗蓝牙协议(BLE)等多种接入方式。

(2)核心服务层。核心服务层由核心数据、命令、元数据、注册表和配置4个微服务组件组成。核心数据微服务存储和管理来自南侧设备的数据、元数据微服务存储和管理设备的元数据。命令微服务将定义在设备文件的操作命令转换成通用的API,提供给用户以监测控制该设备。注册表和配置微服务存储设备服务的相关信息。

(3)支持服务层。支持服务层提供边缘分析和智能服务,以规则引擎微服务为例,允许用户设定一些规则,当检测到数据满足规则要求时,将触发一个特定的操作。例如,规则引擎可监测控制温度传感器,当检测到温度低于25℃时,触发对空调的关闭操作。

(4)导出服务层。导出服务层用于将数据传输至云计算中心,由客户端注册和分发等微服务组件组成。前者记录已注册的后端系统的相关信息,后者将对应数据从核心服务层导出至指定客户端。

(5)系统管理和安全服务。系统管理服务提供安装、升级、启动、停止和监测EdgeX Foundry微服务的功能。安全服务用以保障来自设备的数据和对设备的操作安全。

最新版本的EdgeX Foundry没有为用户自定义应用提供计算框架,用户可以将应用部署在网络边缘,将该应用注册为导出客户端,进而将来自设备的数据导出至应用来处理。EdgeX Foundry的设计满足硬件和操作系统无关性,并采用微服务架构。EdgeX Foundry中的所有微服务能够以容器的形式运行于各种操作系统,且支持动态增加或减少功能,具有可扩展性。EdgeX Foundry的主要系统特点是为每个接入的设备提供通用的Restful API以操控该设备,便于大规模地监测控制物联网设备,满足物联网应用的需求。EdgeX Foundry的应用领域主要在工业物联网,如智能工厂、智能交通等场景,以及其他需要接入多种传感器和设备的场景。

## 2.2 Apache Edgent

Apache Edgent是一个开源的编程模型和微内核风格的运行时,它可以被嵌入到边缘设备上,用于提供对连续数据流的本地实时分析<sup>[5]</sup>。Edgent解决的问题是如何对来自边缘设备的数据进行高效的分析

处理。为加速边缘计算应用在数据分析处理上的开发过程,Edgent 提供一个开发模型和一套 API 用于实现数据的整个分析处理流程。

基于 Java 等的开发环境,Edgent 应用的开发模型如图 3 所示。

该模型由提供者、拓扑、数据流、数据流的分析处理、后端系统 5 个组件组成。

(1)提供者。一个提供者对象包含了有关 Edgent 应用程序的运行方式和位置信息,并具有创建和执行拓扑的功能。

(2)拓扑。拓扑是一个容器,描述了数据流的来源和如何更改数据流的数据。数据的输入、处理和导出至云的过程都记录在拓扑中。

(3)数据流。Edgent 提供了多种连接器以不同方式接入数据源,比如支持消息队列遥测传输(MQTT)、超文本传输协议(HTTP)和串口协议等,用户还可以添加自定义代码以控制传感器或设备的数据输入。

(4)数据流的分析处理。Edgent

提供一系列功能性的 API 以实现数据流的过滤、分裂、变换等。

(5)后端系统。由于边缘设备的计算资源稀缺,Edgent 应用程序无法支撑复杂的分析任务。用户可以使用连接器,通过 MQTT 和 Apache Kafka 方式连接至后端系统,或者连接至 IBM Watson IoT 平台进一步对数据做处理。

Edgent 应用可部署于运行 Java 虚拟机的边缘设备中,实时分析来自传感器和设备的数据,减少了上传至后端系统如云数据中心的数据量,并降低了传输成本。Edgent 的主要系统特点是提供了一套丰富的数据处理 API,切合物联网应用中数据处理的实际需求,降低应用的开发难度并加速开发过程。Edgent 的主要应用领域是物联网,此外,它还可以被用于分析日志、文本等类型的数据。

### 3 面向边缘云服务的边缘计算开源平台

网络运营商的网络边缘,如蜂

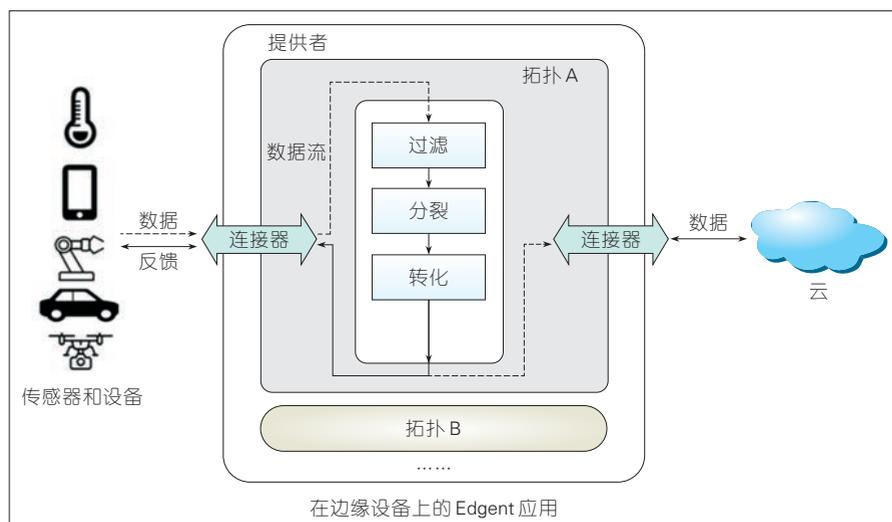
窝网络基站、中央端局和在网络边缘的小型数据中心等,是用户接入网络的地方,其计算、存储和网络资源也可用以部署边缘计算应用。面向边缘云服务的边缘计算平台着眼于优化或重建网络边缘的基础设施以实现在网络边缘构建数据中心,并提供类似云中心的服务。代表性的平台有开放网络基金会(ONF)的 CORD 项目<sup>[6]</sup>和 Linux 基金会的 Akraino Edge Stack 项目<sup>[7]</sup>。

#### 3.1 CORD

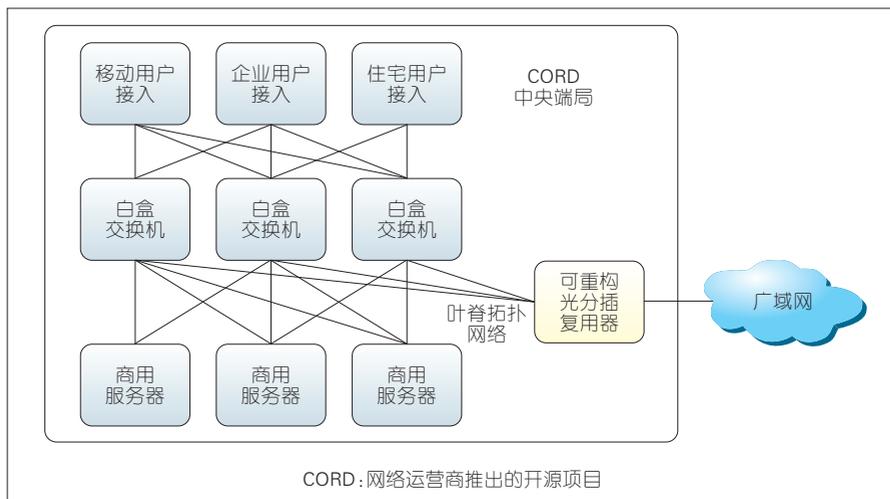
CORD 是为网络运营商推出的开源项目,旨在利用软件定义网络(SDN)、网络功能虚拟化(NFV)和云计算技术重构现有的网络边缘基础设施,并将其打造成可灵活地提供计算和网络服务的数据中心。现有网络边缘基础设施构建于由电信设备供应商提供的封闭式专用的硬件系统,不具备可扩展性,无法动态调整基础设施的规模,导致资源的低效利用。CORD 计划利用商用硬件和开源软件打造可扩展的边缘网络基础设施,并实现灵活的服务提供平台,支持用户的自定义应用。

图 4 为 CORD 的硬件架构图<sup>[6]</sup>,CORD 利用商用服务器和白盒交换机提供计算、存储和网络资源,并将网络构建为叶脊拓扑架构以支持横向网络的通信带宽需求。此外,CORD 使用专用接入硬件将移动、企业和住宅用户接入网络中。

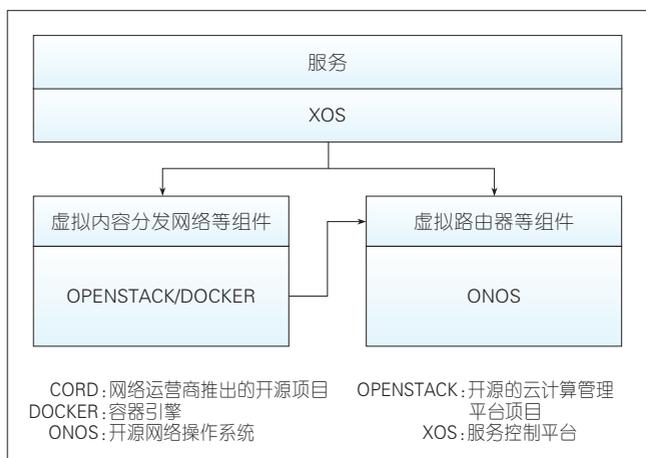
CORD 的软件架构<sup>[6]</sup>如图 5 所示,云平台管理项目 OpenStack 用以管理计算和存储资源,创建和配置虚拟机以及提供基础设施即服务



▲图 3 Edgent 应用的开发模型



▲图4 CORD的硬件架构



◀图5 CORD的软件架构

(IaaS)功能。开源网络操作系统(ONOS)为网络提供控制平面,用于管理网络组件如白盒交换网络结构等,并提供通信服务。容器引擎Docker使用容器技术来实例化提供给用户的服务。服务控制平台XOS用于整合上述软件,以组装、控制和组合服务。

根据用户类型和使用案例的不同,CORD可被具体实现为M-CORD、R-CORD和E-CORD。以M-CORD为例,M-CORD面向无线网络(特别是5G网络),使用蜂窝网络将用户接入。M-CORD基于

NFV和云计算技术将蜂窝网络功能进行分解和虚拟化,实现网络功能的动态扩展同时增强资源利用率。在此基础上,M-CORD支持多接入边缘服务,为用户提供定制服务和差异化体验质量(QoE)。此能力切合具备移动性的边缘计算应用的需求,能通过无线网络为手机、无人车和无人机等移动设备的边缘计算应用就近提供强大的计算能力。R-CORD和E-CORD同样可以在网络边缘支持住宅用户或企业用户的边缘计算应用,如VR和AR应用等,以获得更快的响应时间和更好

的服务体验。

对于用户而言,CORD在运营商网络边缘提供的边缘云服务,具有与云计算相同的优点,即无需用户提供计算资源和搭建计算平台,降低了硬件和管理成本。此外,有线和无线网络的广泛分布使用户提交边缘计算应用不受地理位置的影响。目前,运营商正研究如何部署CORD,中国联通发起成立了CORD产业联盟,推动CORD项目的落地。

### 3.2 Akraino Edge Stack

Akraino Edge Stack是一个面向高性能边缘云服务的开源项目,并为边缘基础设施提供整体的解决方案。Akraino Edge Stack致力于发展一套开源软件栈,用于优化边缘基础设施的网络构建和管理方式,以满足边缘计算云服务的要求,例如高性能、低延迟和可扩展性等。

Akraino Edge Stack项目涉及的范围从基础设施延伸至边缘计算应用,其范围可以划分为3个层面<sup>[7]</sup>。在最上面的应用层面,Akraino Edge Stack致力于打造边缘计算应用程序的生态系统以促进应用程序的开发。中间层面着眼于开发中间件和框架以支持上面层次的边缘计算应用。在这个层面中,Akraino Edge Stack将开发API和框架以接入现有互补性的开源边缘计算项目,例如上述的面向物联网的互操作性框架EdgeX Foundry,最大化利用开源社区的现有成果。在最下面的基础设施层面中,Akraino Edge Stack将提供一套开源软件栈用于

优化基础设施。此外, Akraino Edge Stack 为每种使用案例提供蓝图以构建一个边缘计算平台。每个蓝图涵盖上述3个层次的声明性配置,其中包括对硬件、各层面的支撑软件、管理工具和交付点等的声明。

Akraino Edge Stack 基于使用案例提供边缘云服务,可部署于电信运营商的塔楼、中央端局或线缆中心等。其应用领域包括边缘视频处理、智能城市、智能交通等。

#### 4 面向云边融合的边缘计算开源平台

云计算服务提供商是边缘计算的重要推动者之一,基于“云边融合”的理念,致力于将云服务能力拓展至网络边缘。目前,亚马逊公司推出了 AWS Greengrass<sup>[8]</sup>,微软公司推出了 Azure IoT Edge<sup>[9]</sup>,阿里云公司推出了物联网边缘计算平台 Link IoT Edge<sup>[10]</sup>,微软公司在2018年宣布将 Azure IoT Edge 开源。

Azure IoT Edge 是一种混合云和边缘的边缘计算框架,旨在将云功能拓展至如路由器和交换机等具备计算能力的边缘设备上,以获得更低的处理时延和实时反馈。Azure IoT Edge 运行于边缘设备上,但使用与云上的 Azure IoT 服务相同的编程模型;因此,用户在开发应用的过程中除对计算能力的考量外,无需考虑边缘设备上部署环境的差异,可以将云上原有的应用迁移至边缘设备上运行。

如图6所示, Azure IoT Edge 由 IoT Edge 模块、IoT Edge 运行时和 IoT Edge 云界面组成,前两者运行

在边缘设备上,后者则是一个在 Azure 云上提供服务的管理界面。

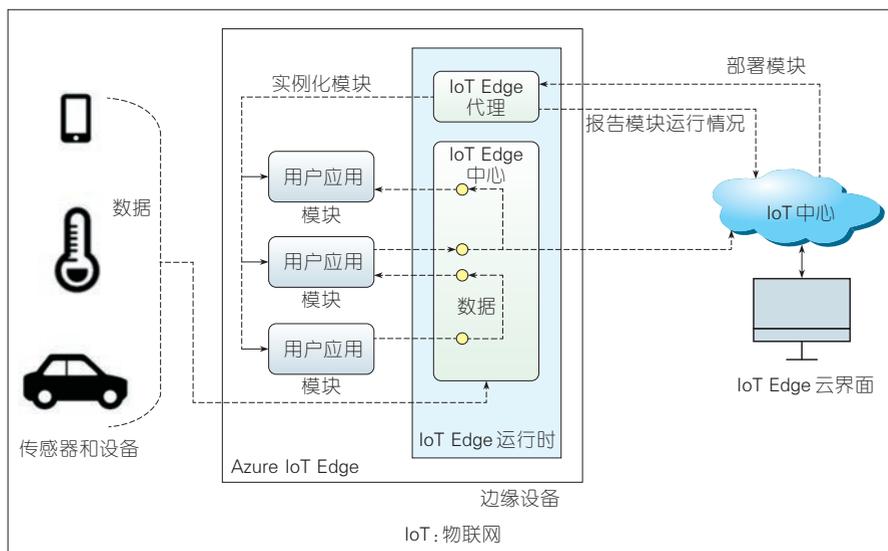
(1)IoT Edge 模块。IoT Edge 模块对应于用户的边缘计算应用程序。一个模块镜像即一个 Docker 镜像,模块里包含用户的应用代码,而一个模块实例就是一个运行着对应的模块镜像的 Docker 容器。基于容器技术, IoT Edge 具备可扩展性,用户可动态添加或删除边缘计算应用。由于相同的编程模型, Azure 机器学习和 Azure 数据流分析等 Azure 云服务也可以部署到 IoT Edge 模块,此特性便于在网络边缘部署复杂的人工智能应用,加快了开发过程。

(2)IoT Edge 运行时。IoT Edge 运行时由 IoT Edge 中心和 IoT Edge 代理2个组件构成,前者负责通信功能,后者负责部署和管理 IoT Edge 模块,并监测控制模块的运行。IoT 中心是在 Azure 云上的消息管理中心, IoT Edge 中心与 IoT 中心连接并充当其代理。IoT Edge 中

心通过 MQTT、高级消息队列协议 (AMQP) 和 HTTPS 协议获取来自传感器和设备的数据,实现设备接入的功能。此外, IoT Edge 中心作为消息中转站,连接 IoT Edge 模块之间的消息通信。IoT Edge 代理从 IoT Hub 接收 IoT Edge 模块的部署信息,实例化该模块,并保证该模块的正常运行,如对故障模块进行重启,并将各模块的运行状态报告至 IoT 中心。

(3)IoT 云界面。云界面提供了设备管理的功能。用户通过云界面进行添加设备,部署应用和监测控制设备等操作,为用户大规模部署边缘计算应用提供了方便。

Azure IoT Edge 的主要系统特点是有强大的 Azure 云服务的支持,尤其是人工智能和数据分析服务的支持。Azure IoT Edge 具有广阔的应用领域,除了物联网场景,原有在云上运行的应用也可以根据需求迁移至网络边缘上运行。目前 Azure IoT Edge 已有智能工厂、智能



▲图6 Azure IoT Edge的架构图

灌溉系统等使用案例。

## 5 构建边缘计算平台的开源软件

一些开源软件能为边缘计算平台提供组件功能,具体包括:

(1)网络管理。开源网络自动化平台(ONAP)<sup>[11]</sup>是一个为物理或虚拟网络功能的自动化和实时策略驱动化编排提供平台的开源项目,在边缘云中可用于定制网络服务。Linux 基金会于2018年将ONAP与其他5个功能类似的开源项目合并成 LF Networking Fund<sup>[12]</sup>项目。ONOS<sup>[13]</sup>是一个开源网络操作系统,它提供一个控制平面用于管理网络组件,运行软件程序和模块,并为终端主机提供通信服务。目前,ONAP应用于 Akraino Edge Stack 项目中,ONOS应用于CORD项目中。

(2)容器技术。Docker作为一个开源的应用容器引擎,被多个边缘计算平台使用以提供灵活的应用部署方式。Kubernetes<sup>[14]</sup>是一个对容器化应用进行自动化部署、扩展和管理的开源项目,可用于边缘计算平台中以提供可靠和可扩展的容器编排。

(3)云平台。OpenStack是一个开源的云管理管理平台项目,可用于构建边缘云,管理边缘云基础设施的计算、存储和网络资源。StarlingX<sup>[15]</sup>是一个用于构建分布式边缘云的开源项目,提供一套完整的云基础架构软件栈,现应用于 Akraino Edge Stack 项目中。

(4)人工智能技术。引入人工智能技术是边缘计算的发展趋势之

一。Acumos 人工智能(AI)<sup>[16]</sup>是一个用于训练、部署和共享 AI 应用程序的开源框架,可用于边缘云中以提高 AI 边缘计算应用的快速开发和部署。Acumos AI 目前已用于 Akraino Edge Stack 项目中。

## 6 结束语

边缘计算平台为基于边缘计算的应用提供了一种降低延迟、提高数据处理实时性的架构和软件栈支撑。目前,边缘计算平台仍处于快速发展时期,本文中我们对其未来的发展趋势进行如下分析:

(1)面向特定应用场景的专用性边缘计算平台将迅速发展。目前,智能家居、智能工厂和智慧城市都是受关注的应用场景,如何系统化支持这些场景的功能需求是边缘计算平台的发展趋势之一。

(2)边缘计算平台的数据安全与隐私保护问题将变得更加突出。边缘设备相互协作时的数据安全如何保证,边缘平台如何避免暴露用户的位置隐私信息、尤其是边缘计算设备的计算资源,存储能力有限,使得边缘计算平台的安全防御具有更高的挑战性。

(3)边缘计算平台将结合更多人工智能技术。目前,以深度学习为代表的人工智能技术发展十分迅速,但“智能边缘”的研究才刚刚起步,未来的边缘计算平台将引入人工智能技术,实现边缘数据的智能处理。

### 参考文献

- [1] SHI W, CAO J, ZHANG Q, et al. Edge Computing: Vision and Challenges[J]. IEEE Internet of Things Journal, 2016, 3(5):637-646. DOI: 10.1109/JIOT.2016.2579198

- [2] 赵梓铭, 刘芳, 蔡志平, 等. 边缘计算: 平台、应用与挑战[J]. 计算机研究与发展, 2018, 55(2): 327-337. DOI: 10.7544/j.issn1000-1239.2018.20170228
- [3] 施巍松, 刘芳, 孙辉, 等. 边缘计算[IM]. 北京: 科学出版社, 2018:3-6
- [4] EdgeX Foundry[EB/OL]. [2018-12-30]. <https://www.edgexfoundry.org>
- [5] Apache Edgent[EB/OL]. [2018-12-30]. <http://edgent.apache.org>
- [6] CORD[EB/OL]. [2018-12-30]. <https://www.opennetworking.org/cord/>
- [7] Akraino Edge Stack[EB/OL]. [2019-02-22] [2018-12-30]. <https://wiki.akraino.org>
- [8] AWS Greengrass[EB/OL]. [2018-12-30]. <https://aws.amazon.com/greengrass/>
- [9] Azure IoT Edge[EB/OL]. [2018-12-30]. <https://azure.microsoft.com/zh-cn/services/iot-edge/>
- [10] 物联网边缘计算[EB/OL]. [2018-12-30]. <https://www.aliyun.com/product/loedge?spm=5176.cnioledge.loT.7.483f61efVjFy39>
- [11] ONAP[EB/OL]. [2018-12-30]. <https://www.onap.org/>
- [12] LF Networking Fund[EB/OL]. [2018-12-30]. <https://www.lfnetworking.org/>
- [13] ONOS[EB/OL]. [2018-12-30]. <https://onosproject.org/>
- [14] Kubernetes[EB/OL]. [2018-12-30]. <https://kubernetes.io/>
- [15] StarlingX[EB/OL]. [2018-12-30]. <https://www.starlingx.io/>
- [16] Acumos AI[EB/OL]. [2018-12-30]. <https://www.acumos.org/>

### 作者简介



梁家越, 中山大学数据科学与计算机学院在读硕士研究生; 研究方向为边缘计算。



刘斌, 中国科学院电子学研究所读博士生; 研究方向为边缘计算。



刘芳, 中山大学数据科学与计算机学院副教授; 研究方向为计算机系统结构与存储技术; 先后主持和参加国家自然科学基金、科技部重点研发、“863”计划等项目10余项; 获中山大学“百人计划”中青年杰出人才, 并获军队科技进步二等奖1次; 发表学术论文70余篇, 授权国家发明专利10余项。

# 边缘存储的发展现状 与挑战

## A Survey on the Storage Issues in Edge Computing

刘铎/LIU Duo, 杨涓/YANG Juan, 谭玉娟/TAN Yujuan

(重庆大学, 重庆 400044)  
(Chongqing University, Chongqing 400044, China)



**摘要:** 边缘存储是支撑边缘计算的核心存储技术。与云存储不同,边缘存储将数据从远距离的云服务器端迁移到离数据更近的边缘存储设备端,具有更低的网络通信开销、交互延迟和带宽成本,能为边缘计算提供实时可靠的数据存储和访问。阐述了边缘存储的概念,描述了边缘存储的典型应用领域,并详细分析了与边缘存储相关的技术研究现状与挑战。

**关键词:** 边缘计算;边缘存储;云存储;分布式存储架构

**Abstract:** Edge storage is the core storage technology to support edge computing. Unlike cloud storage, edge storage stores data in the vicinity of data generation devices, reducing network communication overhead and bandwidth costs, providing real-time data access and high reliability. In this paper, the concept of edge storage is introduced, and the typical application areas, as well as the research trends and challenges of edge storage are analyzed.

**Key words:** edge computing; edge storage; cloud storage; distributed storage architecture

DOI: 10.12142/ZTETJ.201903003

网络出版地址: <http://kns.cnki.net/kcms/detail/34.1228.TN.20190529.1739.006.html>

网络出版日期: 2019-05-29

收稿日期: 2018-12-29

随着物联网技术<sup>[1]</sup>与5G网络的快速发展,网络边缘设备急剧增加,其所产生的数据以爆炸式增长。根据思科可视化网络指数报告<sup>[2]</sup>,到2021年,全球IP数据流量将达到3.3 ZB,其中边缘设备产生的移动数据量占全球数据流量的63%,较2016年增长6.7倍,边缘设备将成为数据生产的主流。高速增长的数据对边缘存储系统的容量、性能、功耗提出了严峻挑战。另一方面,随着深度学习<sup>[3]</sup>和增强现实(AR)技术的广泛应用,边缘设备呈现出智能化趋势,其数据处理需

要具备实时性。而传统云计算模型采用集中式管理<sup>[4]</sup>,该模式需要数据跨越地理位置限制,具有极大的数据传输延迟及网络波动可能性,难以满足边缘应用的实时性需求。

为解决上述问题,边缘存储——一种面向边缘大数据存储的新型分布式存储架构被提出。边缘存储将数据分散存储在邻近的边缘存储设备或数据中心,大幅度缩短了数据产生、计算、存储之间的物理距离,为边缘计算提供高速低延迟的数据访问。边缘计算<sup>[5]</sup>是一种新型分布式计算模型,通过将传统云计

算架构中的部分任务下沉到智能终端设备或边缘计算节点执行,提供实时的数据计算服务。边缘存储为边缘计算的发展提供了较为可靠的支撑。

边缘存储对边缘计算的高效数据存储主要体现在3个方面:第一,边缘存储可提供数据预取和缓存服务,以克服云存储远距离数据传输造成的高延迟、网络依赖等问题;第二,边缘存储可提供邻近边缘终端的分布式数据存储服务,借助数据去重和近似存储技术<sup>[6]</sup>,缓解云数据中心的存储和带宽压力,降低数

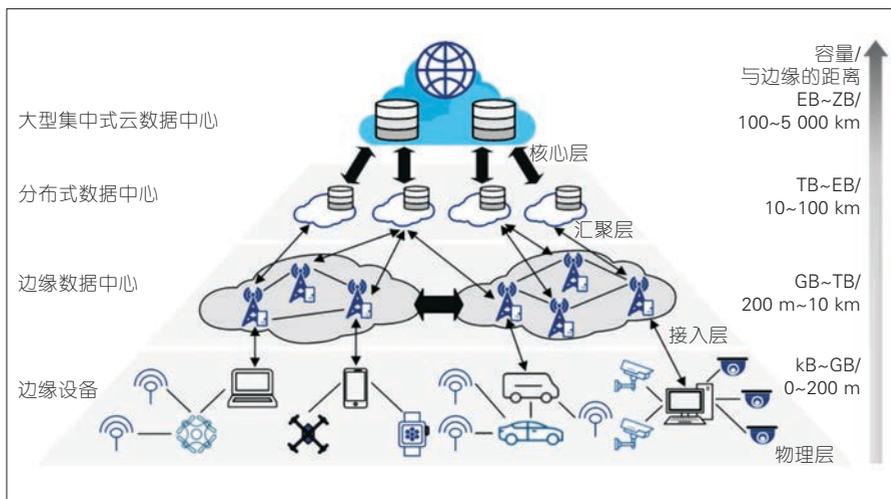
据遭受网络攻击的风险;第三,边缘存储能够与云存储协同提供存储服务,支持边缘计算任务在边缘终端和云端的协同处理,促进边缘计算与云计算的融合。本文中,我们将首先介绍边缘存储的概念,比较边缘存储与云存储的差异;然后介绍边缘存储的典型应用场景,包括工业物联网、车联网、智慧城市、无人机等;最后分析边缘存储相关的技术研究现状与挑战。

## 1 边缘存储的概念

边缘存储是边缘计算的延伸概念,主要为边缘计算提供实时可靠的数据存储与访问。目前工业界和学术界对边缘存储还未进行详细定义。本章节中,我们将从边缘存储的层次结构和特点阐述边缘存储的概念。

### 1.1 边缘存储的层次结构

边缘存储由边缘设备、边缘数据中心、分布式数据中心3层结构组成,如图1所示。顶层为分布式数据中心,部署在距离集中式云较远但互联网用户数量多的城市或地区,为用户提供城域EB级数据存储服务,也称作分布式云,通常与大型集中式云数据中心协同执行存储任务;中间层为边缘数据中心(EDC),也称作边缘云,通常部署在蜂窝基站和人群密集处,为区域内提供TB级实时存储服务,多个小型物理数据中心通过软件定义网络(SDN)可组合成一个逻辑数据中心;底层由数量庞大的边缘设备组成,涵盖桌面电脑、智能手机、传感器、物联网



▲图1 边缘存储的层次化结构

(IoT)网关、传感网执行器以及智能路边单元等多种设备,设备之间可通过无线接入技术相互连接组成边缘存储网络。

### 1.2 边缘存储的特点和优势

不同于集中式云存储服务,边缘存储将数据存储从远距离的云服务器端,迁移到离数据更近的边缘存储设备或边缘数据中心就近存储,具有更低的网络通信开销、交互延迟和带宽成本,更高的自适应能力与可扩展性。如表1所示,与云存储相比,边缘存储具有如下特点和优势:

(1)边缘存储设备具备地理分布式特性。

边缘存储设备和边缘数据中心在地理上是分布式的。大量地理位

置分散的边缘存储设备可借助Wi-Fi、蓝牙、Zigbee等无线接入技术,与相邻的存储设备或边缘数据中心构成分布式存储网络。这种地理分布式结构使数据能够及时地就近存储,为边缘计算关键任务的实时性数据存储和访问提供了保障。而云数据中心在地理上是集中式的,远距离的传输延迟使得大量边缘设备的数据存储和处理需求无法被及时处理;拥塞的网络、高延迟的服务等都将导致服务质量(QoS)的急剧下降。

(2)边缘存储介质和系统具备异构性。

边缘存储具有异构存储的特点,主要体现为水平边缘异构和垂直多层次异构。水平边缘异构是指不同种类的边缘终端设备通常采用

▲表1 边缘存储与云存储的比较

|      | 云存储特点         | 边缘存储特点         | 边缘存储优势          |
|------|---------------|----------------|-----------------|
| 存储设备 | 集中式           | 地理分布式          | 低延迟的实时数据处理      |
| 存储架构 | 部分支持内部部署(私有云) | 支持内部部署         | 更高的安全性和隐私性      |
| 存储系统 | 同构            | 水平边缘异构和层次性垂直异构 | 多层次的数据缓存        |
| 存储数据 | 与数据产生的地理位置无关  | 位置感知           | 更低的网络负载、更快的数据处理 |

不同的存储介质,或者基于同一种存储介质采用不同的存储系统软件进行数据存储。该特性使得边缘存储能利用大量的异构存储介质和存储系统就近快速地存储各类边缘数据。垂直多层次异构是指根据距离大型集中式云数据中心的远近,边缘存储可分为3个层次:边缘设备、边缘数据中心、分布式数据中心,不同层次对应不同的存储系统。该特性使不同层次的存储系统能够相互协作,通过多层次、多级别的数据缓存和预取策略优化边缘数据的存储和访问。

(3)边缘存储架构具备支持内部部署的特性。

边缘存储架构支持在边缘端设备内部部署存储系统,与外部网络进行隔离。独立存储数据于内部部署的边缘存储系统,具有以下优势:能够为边缘计算任务提供高速的本地数据资源访问,满足边缘应用的实时性需求;能够在本地最大限度地控制访问内部存储设备,监测控制数据存储的位置,实时调整机密数据的冗余策略;能面向数据源对数据进行加密或其他预处理,增强数据安全性。

(4)边缘存储数据具备位置感知的特性。

边缘存储就近存储数据,数据分布与地理位置紧密相关,具备地理位置的强感知特性。依托该特性,边缘计算任务在处理数据时,无须查询整个存储网络定位数据,极大地减少主干网络的流量负载。同时,边缘计算任务可以和所需数据在地理位置上近距离绑定,减少数

据在网络上的传输延迟,加快数据的处理速度,为大数据分析平台<sup>[7]</sup>提供更好的底层支持。此外,通过对边缘存储数据的地理分布情况进行统计和分析,应用服务提供商可以联系移动用户与感兴趣的企业和事件,提升用户服务质量。

## 2 边缘计算的典型应用领域

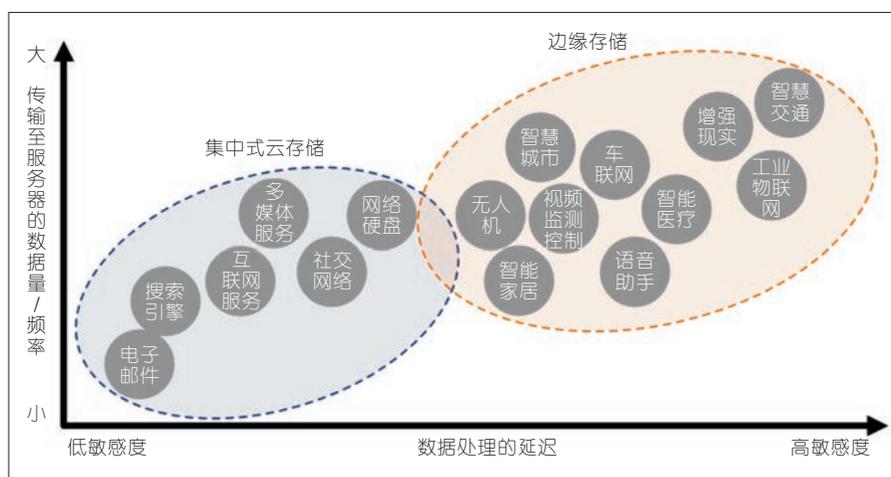
与云存储不同,边缘存储适用于与物联网设备紧密相关、数据传输延迟敏感、数据交互次数多、数据传输量大的物联网应用,如工业互联网、车联网、智慧城市、无人机、视频监控控制等。而云存储则适用于延迟敏感度稍低的互联网应用,如搜索引擎、社交网络、网络硬盘等。边缘存储与云存储应用领域比较如图2所示。本章节中,我们将对边缘存储的几个典型应用领域进行简要介绍。

### 2.1 工业互联网

工业互联网(IIoT)由工业系统各种元素和处理其生成数据的高级分析平台组成。机器对机器

(M2M)通信会产生大量工业数据,云端存储模式不能满足工业物联网的需求,而法规禁止将工业数据传输到其他或公司计算架构之外的云服务器;因此,需要本地实时数据处理支持。

边缘存储支持内部部署,在本地和云之间架起桥梁,实现现场层的高频数据交换,这是IIoT的重要支撑技术。工业物联网边缘存储平台通常由后端的边缘存储管理系统、存储设备以及边缘应用程序组成,提供设备级数据处理,安全地将高度完善的分析技术和边缘智能引入制造领域,使自动化设备得到进一步扩展。例如,西门子公司发布的Industrial Edge产品,其收集自动化组件数据并存储在本地,提供本地数据高速读取和实时处理功能,预处理后的数据转存至云端,云端将执行计算密集型神经网络模型训练,并将改进后的模型部署在边缘端。由于本地存储及预处理了大量数据,只需将最终相关数据传输到云或IT基础架构,降低了数据泄露的风险,减少了内存占用和传输成



▲图2 边缘存储与云存储应用领域比较

本,为工业生产提供持续稳定的生产流程和状态监测控制,显著提高生产力。

## 2.2 车联网

车联网是由联网汽车与交通信息系统互联形成的巨大交互网络,车联网技术允许汽车与车辆内外的其他设备共享互联网访问<sup>[8]</sup>。由于车辆通常处于高速移动状态,联网状态不稳定,边缘存储架构在车联网中具有显著的优势。自动驾驶汽车通常通过路测单元(RSU)实现网络接入和数据处理任务,并依赖于车联网技术采集路边信息、接入边缘计算服务。通过RSU提供边缘存储支持,数据的存放更接近数据源,原始数据进一步丰富,车辆动作、决策的相关信息能快速发送到适当的目的地。HU Z.等人提出了一种基于多目标拍卖的机制,解决RSU缓存中的多服务提供商共存问题<sup>[9]</sup>,仿真结果证明RSU边缘数据缓存能够加速车辆的下载任务。

## 2.3 智慧城市

智慧城市使用信息和通信技术以及物联网设备来提高城市服务质量和运营效率<sup>[10]</sup>,改善公民生活水平,降低能耗,其数据大多来自物联网传感器和网络摄像机。目前,视频监控控制产生的数据正在快速增长。据思科可视化网络指数<sup>[2]</sup>统计,2016年网络视频监控控制流量平均每月883 PB,相比2015年增加72%。尽管云存储架构具备丰富的存储资源,但实时传输视频监控控制数据至云服务器将对网络造成巨

大负担,采用去中心化边缘存储技术可以改善这一情况。传感器收集的数据无须上传至远端的云服务器;而是分散存储在各个边缘存储设备中,为智慧城市提供廉价的存储空间和更高的数据安全保障。此外,位置感知的存储模式为智慧城市的决策制订与实施提供额外数据支持,帮助优化城市功能,推动经济增长,提高市民生活质量,例如水质监测系统传感器信息能够用于污染源的检测和溯源。

## 2.4 无人机

无人机已由军用迈向商用,应用范围也越来越广<sup>[11]</sup>,其中飞行数据采集系统是其应用。该系统通常需要操作员手动完成飞行控制、数据分析以及任务的规划和设计等多个操作,过程繁琐且不利于业务扩展;因此可以使自动化飞行数据采集系统与边缘数据中心协作,将无人机的控制转移到边缘数据中心,通过无人机软件控制平台自动化完成数据收集与分析工作。

对于无人机自动化飞行数据采集系统,边缘存储的意义在于取代手动数据拷贝流程,数据可以实时存储在边缘数据中心(EDC)中并立即执行数据分析算法,控制平台可根据分析结果灵活调整任务流程。市场上已经有公司开展此类研究,无人机自动化飞行数据采集公司Hanger与EDC服务提供商Vapor IO合作,开展无人机与EDC的协作。通过EDC提供的存储与网络支持,无人机能直接将建筑、桥梁、和输送管道的视频传至EDC进行快速

处理和分析,操作员无须手动操作。无人机与EDC的协作,使得飞行数据采集系统更加灵活、更加自动化。

## 3 研究现状与挑战

边缘存储技术的研究目前还处于萌芽阶段<sup>[12]</sup>,本章节中,我们将从边缘存储的介质和设备、边缘分布式存储架构,以及云-边协同3个方面介绍相关技术研究现状与挑战。

### 3.1 边缘存储介质和设备

#### (1)边缘存储介质。

边缘存储终端多使用非易失性存储介质进行数据存储。目前主流的非易失性存储介质为闪存(FLASH)<sup>[13]</sup>,根据实现方法的不同,又分为NOR闪存与NAND闪存。在NOR闪存的内部电路中,各存储器单元并联连接,支持快速的数据随机访问。其适合于容量要求低、速度要求高、数据只读的应用,如存储边缘设备固定代码的基本输入输出系统(BIOS)应用。NAND闪存内部有一系列串联的存储晶体管,相较于NOR闪存,其存储单元尺寸更小、存储密度更大。NAND闪存通过对数据块的编程可以实现更快的写入和擦除,适合于存储成本低、存储密度高、需要进行快速擦写的应用,其特性正适用于成本低、数据量大的边缘设备。

近年来,随着非易失性存储器(NVM)的快速发展,基于NVM的存储技术研究得到广泛关注,如相变存储器(PCM)<sup>[14]</sup>、阻变式随机访问存储器(ReRAM)<sup>[15]</sup>等。将NVM存

储介质使用于边缘设备也逐渐成为研究热点。NVM存储介质集合了动态随机存取存储器(DRAM)和闪存的优点,具有较高的存储密度,可字节寻址、读延迟低,能持久性存储数据。NVM的成本预计介于DRAM和Flash之间,可作为DRAM内存或Flash固态硬盘(SSD)的替换方案。然而,由于材料的限制,NVM自身也存在一些缺陷。与DRAM相比,NVM的写次数有限,写入延迟更高;而与闪存相比,NVM的存储密度较小。因此,NVM在边缘存储中作为内存与外存的一体化解决方案还有很多技术难点需要克服。

## (2)边缘存储设备。

基于闪存芯片制造的SSD常作为边缘存储设备广泛地应用于各种系统中<sup>[16]</sup>。SSD被广泛使用主要得益于其具有很高的输入/输出(I/O)性能、低功耗和高可靠性。不过,SSD的I/O读写性能具有不对称性,读性能往往优于写性能,在边缘设备进行大量数据写入时性能较低。针对这些问题,硬件厂商在边缘设备内部提供闪存转化层(FTL),FTL主要负责数据读写的地址映射与转换以及垃圾数据块的标记与回收,以提高数据的读写性能,延长寿命,并向文件系统层屏蔽硬件调度细节,仅提供数据读写的逻辑接口。但是,FTL封装在SSD中,无法根据上层应用的数据集特征进行动态调整,软件直管的SSD<sup>[17]</sup>是一种有效解决此问题的思路。在Open-channel SSD中,传统SSD的FTL功能将在主机端以软件的方式实现,

不再作为固件封装在边缘设备中,SSD提供给文件系统硬件调度的物理接口。通过Open-channel SSD,便可以根据上层应用的特征定制其FTL的功能,性能更高,灵活性更大,使边缘存储设备更能满足边缘设备对数据存储的需求。

另外,大数据的趋势正在导致计算范式的变化,尤其是将计算转移到数据的概念,我们称之为近数据处理(NDP)方法<sup>[18]</sup>。NDP是指将计算迁移到离数据更近的地方,减少数据的移动。数据的迁移往往在整个计算过程中占据极大的能耗开销,数据的搬运效率却并不会因为摩尔定律的发展而提高,存算一体化逐渐受到研究人员的关注。CHI P.等人提出PRIME<sup>[19]</sup>架构,即可在内存中实现神经网络的计算。由于PRIME架构极大地减少了数据的搬移,性能和能耗分别改进了2360倍和895倍。将边缘存储与存算一体化紧密结合,依靠设备或介质的计算能力,直接在存储控制器内部执行计算,在边缘设备上可极大减少存储器内部的数据迁移开销,在网络传输过程中也能极大减少带宽开销,这将极大减轻云端的计算压力。

(3)挑战:边缘设备上的数据的抽象。

边缘设备多样化,不同边缘设备上存储数据的标准不尽相同,为了使数据能在多种边缘设备之间交互或在边缘数据中心进行统一分析处理,需要对边缘数据进行抽象。在边缘端进行数据抽象极具挑战性,主要体现在以下3点:其一,边

缘设备的差异性,不同边缘设备制造商可能采取不同的数据格式标准,边缘存储系统需要适配各个设备制造商的数据抽象标准;其二,边缘数据的多样性,例如摄像头捕捉的视频流数据、温度传感器采集的温度变化曲线、移动用户访问文件的频率,以上数据均从边缘端产生并传输至边缘数据中心存储,因此边缘存储系统需要设计不同的存储策略,对视频流数据压缩处理,对温度数据可以采取近似存储处理,对访问频率数据可以备份至云端进行大数据分析;其三,边缘数据预处理的不规范性,边缘数据在传输至边缘数据中心之前可能经过了数据清洗和有损压缩,导致部分信息丢失,无法与其他数据进行整合进行大数据分析,因此亟待制定统一的数据抽象和预处理标准。

## 3.2 边缘分布式存储架构

### 3.2.1 边缘分布式存储系统的2种架构

边缘分布式存储系统管理海量数据和存储设备,需要强大的边缘数据中心或高效的自组织边缘存储集群提供硬件支持。不同于传统的云端存储服务提供者,边缘分布式存储系统在网络拓扑结构中更靠近边缘设备,具有更少的通信开销和更高的服务质量。从组织方式的角度来看,边缘分布式存储系统可以分为中心化与去中心化2种架构。

(1)中心化分布式存储架构。

中心化分布式存储通常采取主/从式架构:主节点具备丰富的计

算和存储资源,负责存储节点的管理、存储任务的调度、数据布局以及数据的一致性维护等;从节点仅具备简单的数据存储功能。最典型的中心化分布式存储技术有分布式文件系统(HDFS)<sup>[20]</sup>,它以流式数据访问模式来存储超大规模的数据文件,NameNode(管理节点)作为主节点管理集群,DataNode(工作节点)作为从节点负责处理数据。中心化分布式存储架构可以应用于边缘数据中心。边缘数据中心类似于云存储数据中心。与云存储数据中心相比,边缘数据中心在地理位置上离边缘设备更近,节点规模更小。边缘设备中的数据可上传至边缘数据中心进行存储和管理,云存储数据中心的数据也可在边缘数据中心进行缓存和预取。

#### (2)去中心化分布式存储架构。

去中心化分布式存储没有中心节点,节点之间具有对等的功能。多个边缘设备之间可以自组织地建立去中心化分布式存储网络。随着边缘设备数量激增,该架构具有很大的潜力。如 Storj Labs 推出了一种去中心化分布式云存储平台 STORJ<sup>[21]</sup>,该平台使用点对点网络连接存储设备,并借助以太坊区块链技术发行 STORJ 代币,用户可以在该平台选择出租闲置存储资源获取代币,或使用代币购买存储空间。这种去中心化的分布式存储架构能将很多闲置的存储资源充分利用起来,以非常低廉的维护和管理成本为边缘端提供存储服务。此外,这种结构使数据在边缘端就近存储,更容易满足边缘计算任务的

实时性数据处理需求,比传统的云存储服务更加经济高效。

### 3.3.2 融合中心化和去中心化的分布式存储架构

边缘数据中心的中心化分布式存储能更好地保证服务质量和数据的一致性,适合需要高可靠性和高协作性的应用任务。边缘设备端自组织的去中心化分布式存储具备造价成本低、可灵活部署等特性。随着边缘设备的更新或迁移,可依据 2 种架构各自的优势,融合 2 种架构,提高边缘存储网络的可靠性和普适性。而两者的融合需要解决中心化存储与去中心化存储的无缝切换问题。现今的分布式文件系统无法做到中心化到去中心化的无缝切换,这将严重影响边缘存储系统的可靠性。例如,当采用中心化设计的边缘存储系统断开与中心服务器连接后,大部分存储操作将受到限制。若采用去中心化设计,边缘设备无法利用中心服务器维持数据一致性,不适用多人协作的应用场景。中心化存储系统和去中心化存储系统之间如何互联互通、相互融合是边缘存储技术中极具挑战性的问题。

### 3.3 “云-边”协同

在现有的云存储和边缘存储架构上,为了更好地满足边缘应用计算和大数据处理的需求,需要云端和边缘端协同执行存储任务,以提高数据处理的实时性、可靠性和安全性。本节中,我们将从边缘数据的预取和缓存,以及云-边协同调

度 2 方面阐述云端和边缘端如何协同执行存储任务的现状与挑战。

#### 3.3.1 云-边协同相关现状

##### (1)边缘数据预取和缓存。

由于边缘存储设备离云存储数据中心较远,边缘数据的高效预取和缓存是提高数据访存性能的关键。边缘存储设备访问数据内容之前通常需要将数据从云存储数据中心的服务器下载到边缘服务器中,以降低边缘设备访问数据的时间,这叫做边缘数据预取技术。该技术的关键点在于如何对预先存储的数据进行选择。常用的数据预取算法主要基于访问时间、频率、数据大小、优先级和关联度来建立预取模型。基于访问时间的预取模型是假设上次被访问的数据具有较大可能被再次访问;基于频率的预取模型采用用户访问数据的频率来预测未来的访问情况;基于关联度的预取模型采用机器学习算法挖掘用户兴趣关联规则作为预测依据。常用的数据预取算法可适用于边缘数据的预取。此外,也有些新的边缘数据预取算法被提出。如 HA K. 等人采用了基于应用场景的预取模型<sup>[22]</sup>,根据实际应用场景预测数据请求;U.DROLIA 等人利用马尔可夫模型预测用户可能使用到的神经网络模型并预取这些模型到边缘<sup>[23]</sup>。

边缘数据缓存技术通过缓存历史文件的方式加速用户的访问速度。当用户请求的静态内容没有存储在本地缓存中时,用户向内容分发的网络(CDN)发送请求;CDN 在边缘高速缓存中搜索请求内容<sup>[24]</sup>,

然后将命中的内容传递给用户。通过此技术能减少回程(Backhaul)网络的数据传输开销(如图3所示),降低数据中心的能耗,创造更高的经济效益。边缘缓存技术的关键在于缓存位置的选择,依据与边缘用户的距离分为3种:边缘设备、边缘数据中心和分布式云。将缓存保存在边缘设备能够提供数毫秒级的延迟,但边缘设备存储容量较小并且距离云端远;因此缓存命中率较低,缓存未命中的代价较高。同理,缓存在分布式云能够取得较高的缓存命中率但延迟较高。CHEN B.等人提出一种基于设备到设备(D2D)网络的边缘缓存模型<sup>[25]</sup>,将边缘用户按集群划分,把热文件缓存在各个集群中,此策略可提升4倍D2D缓存网络吞吐量。L.RAMASWAMY等人提出合作边缘缓存网格<sup>[26]</sup>,该网格由多个分布式的边缘缓存云组成,如果缓存未命中,边缘缓存节点能够从临近的缓存云中获取文件数据,无须从云端刷新缓存,减少用户等待时间。

#### (2)云-边协同调度。

如果仅对云端或边缘终端存储体系结构进行优化,忽略其协同与融合,便无法充分发挥两者各自的

优势。对于大规模数据,边缘存储与云存储相互协作才能最大限度发挥云存储地理集中式和边缘终端地理分布式的优势。云-边协同存储技术旨在通过边缘存储与云存储的互补,提供更高效率的存储服务。诸多研究表明,通过对边缘存储资源的有效使用和管理,能够缓解云服务器存储资源紧张<sup>[21-24]</sup>,有效节省能源<sup>[6][27]</sup>,从而提升边缘计算应用性能<sup>[23][25][28]</sup>。F. JALALI通过对云存储应用的能耗进行分析,发现将部分应用迁移到边缘端能够显著减少能源消耗<sup>[27]</sup>。悉尼大学的研究者提出边缘网络和云平台的资源整合框架,该方案设计了用于众包传感器云服务的两级组合模型,抽象了云上传感器数据的时间、空间特性,使用一种基于三维R-树的时空索引技术,用于快速识别合适的传感器云服务,并面向服务质量(QoS)指标为边缘节点选择最优的云服务组合方案<sup>[23]</sup>。阿廷根大学与南京大学的研究者研究了多信道无线干扰环境下移动边缘云计算的多用户计算卸载问题,采用博弈论的方法以分布式方式实现高效的卸载计算,使云平台 and 边缘终端的整体性能达到纳什均衡,从而最大化地利用边

缘数据中心的计算资源<sup>[23]</sup>。上述工作从资源调配的角度研究了云-边协同的存储架构和优化技术。

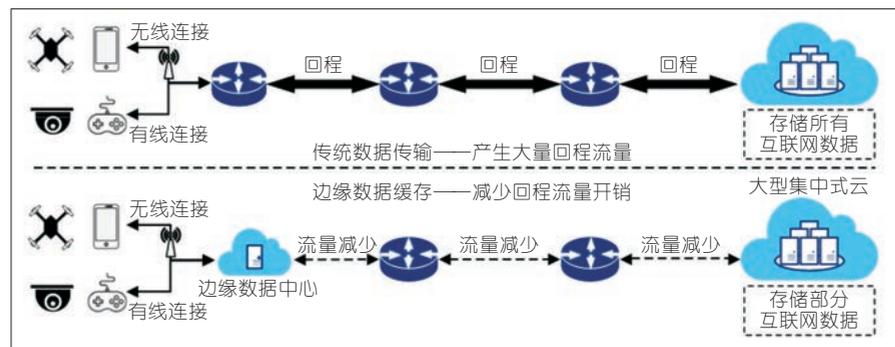
### 3.3.2 云-边协同面临的挑战

(1)高效的数据预取以及缓存策略。

边缘计算任务对低延迟的需求推动了边缘存储的发展,所以边缘数据缓存策略是边缘数据中心设计过程中首要考虑的问题。边缘计算需要适当的边缘预取与缓存技术来加速数据访问速度,并且需要与云服务器协同完成部分数据同步任务。针对延迟容忍度低并且时效性要求较高的应用,边缘数据中心需要动态调整任务的优先级,优先满足延迟敏感应用的存储需求。传统基于CDN采取的缓存策略仅将数据缓存在专用服务器上,通过反向代理的方式提供加速服务,该缓存策略无法结合数据的地理分布信息和区域热度进行调整,不适用于边缘数据缓存的情况。所以重新设计一种适用于边缘存储环境下的数据缓存策略具有重大意义。

#### (2)边缘存储的计费模式。

边缘存储需要合理的计费模式。在云计算架构下的计费由核心网负责,而边缘存储将存储服务分散到边缘数据中心和底层边缘设备中。由于边缘存储网络涉及多个厂家的多种不同的存储设备,设备之间通过协同操作提供边缘存储服务,边缘存储的计费问题十分复杂。在设计出合理的付费模式之前,云服务使用者将惯性维持现有商业行为。因此,探索适合边缘存



▲图3 边缘数据缓存与传统数据传输

储的付费模式对于推动边缘存储商业化非常重要。

## 4 结束语

边缘存储技术目前还处于萌芽阶段。5G时代的到来将我们推向了边缘大数据时代,研究适用于边缘计算架构的边缘存储技术是未来趋势。边缘存储技术支持地理分布式存储架构、内部部署,并能感知基于地理位置的数据热度,能够应对未来日益增长的数据规模、日趋严格的数据响应延迟、数据安全和隐私等需求,必将成为实现边缘数据实时分析处理和有效管理的新途径。因此,需要进一步推动物联网、人工智能、智慧城市、微型数据中心等相关技术和应用的推广和发展。

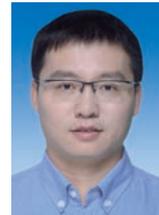
## 致谢

感谢重庆大学陈成彰博士在文章修改过程中提供的宝贵意见,段莫名、张润宇、吴宇同学也在文献收集、整理方面提供了大力帮助,谨致谢意!

## 参考文献

- ASHTON K. That 'Internet of Things' Thing [J]. *RFID Journal*, 2009, 22(7): 97–114
- Cisco. Cisco Visual Networking Index: Forecast and Methodology [R]. 2017
- LECUAN Y, BENGIO Y, HINTON G. Deep Learning [J]. *Nature*, 2015, 521(7553): 436–444. DOI:10.1038/nature14539
- ARMBRUST M, STOICA I, ZAHARIA M, et al. A View of Cloud Computing[J]. *Communications of the ACM*, 2010, 53(4): 50. DOI:10.1145/1721654.1721672
- 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
- 吴宇, 杨涓, 刘人萍, 等. 近似存储技术综述[J]. *计算机研究与发展*, 2018, 55(9): 2002–2015
- 郑纬民, 陈文光. 大数据分析平台: 从扩展性优先到性能优先[J]. *中兴通讯技术*, 2016, 22(2): 11–13. DOI:10.3969/j.issn.1009-6868.2016.02.003
- ZHANG S, CHEN J Y, LYU F, et al. Vehicular Communication Networks in the Automated Driving Era[J]. *IEEE Communications Magazine*, 2018, 56(9): 26–32. DOI:10.1109/mcom.2018.1701171
- HU Z, ZHENG Z, WANG T, et al. Poster: Roadside Unit Caching Mechanism for Multi-Service Providers[C]//Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing. USA:ACM, 2015: 387–388
- CARAGLIU A, DEL BO C, NIJKAMP P. Smart Cities in Europe[J]. *Journal of Urban Technology*, 2011, 18(2): 65–82. DOI: 10.1080/10630732.2011.601117
- MOTLAGH N H, BAGAA M, TALEB T. UAV-Based IoT Platform: A Crowd Surveillance Use Case[J]. *IEEE Communications Magazine*, 2017, 55(2): 128–134. DOI: 10.1109/mcom.2017.1600587cm
- KASEY P. Top Trends in the Gartner Hype Cycle for Emerging Technologies[EB/OL]. [2019–12–20].<https://www.gartner.com/smarterwithgartner/top-trends-in-the-gartner-hype-cycle-for-emerging-technologies-2017/>
- BEZ R, CAMERLENGHI E, MODELLI A, et al. Introduction to Flash Memory[J]. *Proceedings of the IEEE*, 2003, 91(4): 489–502. DOI: 10.1109/jproc.2003.811702
- BURR G W, BREITWISCH M J, FRANCESCHINI M, et al. Phase Change Memory Technology[J]. *Journal of Vacuum Science & Technology B*, 2010, 28(2): 223–262. DOI:10.1116/1.3301579
- AKINAGA H, SHIMA H. Resistive Random Access Memory (ReRAM) Based on Metal Oxides[J]. *Proceedings of the IEEE*, 2010, 98(12): 2237–2251. DOI:10.1109/jproc.2010.2070830
- CHEN F, KOUFATY D A, ZHANG X. Understanding Intrinsic Characteristics and System Implications of Flash Memory based Solid State Drives[C]// Proceedings of the Eleventh International Joint Conference on Measurement and Modeling of Computer Systems. USA:ACM,2019, 37(1):181–192
- BJORLING M, GONZALEZ J. LightNVM: The Linux Open-Channel SSD Subsystem[EB/OL]. [2018–12–20].[file:///C:/Users/susuyeye612/AppData/Local/Microsoft/Windows/Temporary%20Internet%20Files/Content.IE5/S7313G2Q/Matias\\_Bj\\_rling\\_\\_\\_Javier\\_Gonzales.pdf](file:///C:/Users/susuyeye612/AppData/Local/Microsoft/Windows/Temporary%20Internet%20Files/Content.IE5/S7313G2Q/Matias_Bj_rling___Javier_Gonzales.pdf)
- BALASUBRAMONIAN R, CHANG J C, MANNING T, et al. Near-Data Processing: Insights from a MICRO-46 Workshop[J]. *IEEE Micro*, 2014, 34(4): 36–42. DOI:10.1109/mm.2014.55
- CHI P, LI S, XU C, et al. Prime: A novel processing-in-memory architecture for neural network computation in rera-m-based main memory[C]//2016 ACM/IEEE 43rd Annual International Symposium on Computer Architecture (ISCA). USA: IEEE, 2016. DOI: 10.1109/ISCA.2016.13
- SHVACHKO K, KUANG H R, RADIA S, et al. The Hadoop Distributed File System[C]//2010 IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST). USA: IEEE, 2010: 1–10. DOI:10.1109/MSST.2010.5496972
- WILKINSON S, BOSHEVSKI T, BRANDOFF J, et al. storj a Peer-to-Peer Cloud Storage Network[EB/OL]. [2019–05–23].<https://storj.io/storj2014.pdf>
- HA K, PILLAI P, LEWIS G, et al. The Impact of Mobile Multimedia Applications on Data Center Consolidation[C]//2013 IEEE International Conference on Cloud Engineering (IC2E). USA: IEEE, 2013: 166–176. DOI:10.1109/IC2E.2013.17
- DROLIA U, GUO K, NARASIMHAN P. Precog: Prefetching for Image Recognition Applications at the Edge[C]//Proceedings of the Second ACM/IEEE Symposium on Edge Computing. USA:ACM, 2017: 17
- NYGREN E, SITARAMAN R K, SUN J. The Akamai Network[J]. *ACM SIGOPS Operating Systems Review*, 2010, 44(3): 2. DOI: 10.1145/1842733.1842736
- CHEN B Q, YANG C Y, WANG G. High-Throughput Opportunistic Cooperative Device-to-Device Communications with Caching[J]. *IEEE Transactions on Vehicular Technology*, 2017, 66(8): 7527–7539. DOI: 10.1109/tvt.2017.2659701
- RAMASWAMY L, LIU L, IYENGAR A. Scalable Delivery of Dynamic Content Using a Cooperative Edge Cache Grid[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2007, 19(5): 614–630. DOI: 10.1109/tkde.2007.1031
- JALALI F. Energy Consumption of Cloud Computing and Fog Computing Applications [EB/OL]. [2018–12–20]. <https://minerva-access.unimelb.edu.au/handle/11343/58849>
- CHEN X, JIAO L, LI W Z, et al. Efficient Multi-User Computation Offloading for Mobile-Edge Cloud Computing[J]. *ACM Transactions on Networking*, 2016, 24(5): 2795–2808. DOI:10.1109/tnet.2015.2487344

## 作者简介



刘铎,重庆大学计算机学院教授、博士生导师;研究方向为智能物联网系统与边缘计算、大数据处理与先进存储;曾主持或作为主研人参与了国家“863”重大项目、重庆市科技攻关重点项目、国家自然科学基金、教育部博士点基金等。



杨涓,重庆大学计算机学院计算机科学与技术专业在读研究生;研究方向为边缘存储、近似存储等。



谭玉娟,重庆大学计算机学院副教授、硕士生导师;研究方向为重复数据删除、新型非易性存储和智能存储;曾主持或作为主研人员参与了国家“863”重大项目、国家自然科学基金、教育部博士点基金等。

# 移动边缘计算： 架构、应用和挑战

## Mobile Edge Computing: Architectures, Applications and Challenges

高志鹏/GAO Zhipeng, 尧聪聪/YAO Congcong, 肖楷乐/XIAO Kaile

(北京邮电大学网络与交换技术国家重点实验室, 北京 100876)  
(State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and  
Telecommunications, Beijing 100876, China)



**摘要:** 通过对移动边缘计算(MEC)的概念、框架、应用场景等详细阐述,认为 MEC 作为一种新的计算范式,一方面可以满足用户设备飞速增长的计算需求,另一方面也可以提高用户体验的质量。另外,MEC 的及时性和即时性等特征为时延敏感应用的发展提供了良好的基础。

**关键词:** MEC; 架构; 5G; 物联网; 车联网; 安全和隐私; 增强现实

**Abstract:** Through detailed descriptions of the concepts, frameworks, and application scenarios of mobile edge computing (MEC) in this paper, we believe that MEC, as a new computing paradigm, can meet the computational demands of user equipment, and improve the quality of user experience. What's more, the timeliness and immediacy of MEC provide a good foundation for the development of delay-sensitive applications.

**Key words:** MEC; architecture; 5G; Internet of things; Internet of vehicles; security and privacy; augmented reality

DOI: 10.12142/ZTETJ.201903004  
网络出版地址: <http://kns.cnki.net/kcms/detail/34.1228.TN.20190529.1506.002.html>

网络出版日期: 2019-05-29  
收稿日期: 2018-12-30

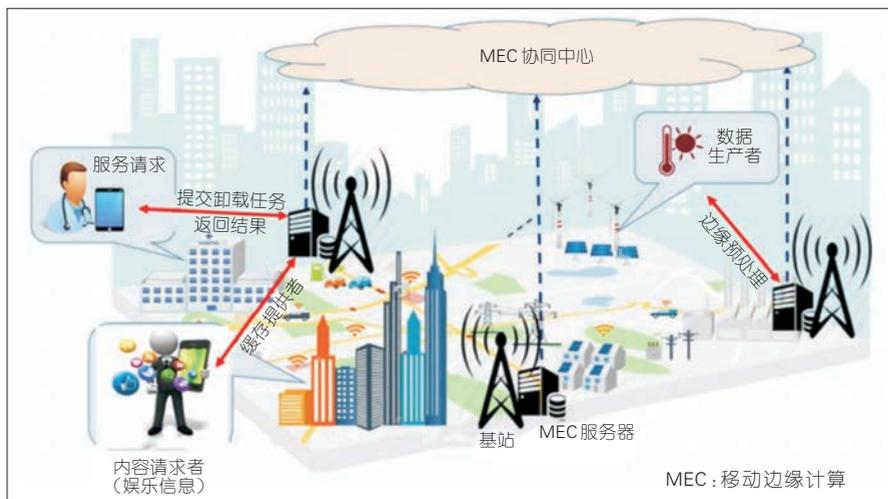
### 1 移动边缘计算的诞生和发展

在过去的十几年中,云计算把计算能力、存储能力和网络管理集中化,其中网络管理指的是对数据中心网、骨干IP网以及蜂窝核心网络的管理。云计算通过云中大量可用的资源对资源受限的终端用户设备提供弹性计算和数据存储功能。云计算推动了许多互联网公司的快速增长,例如软件即服务<sup>[1]</sup>被广泛应用到谷歌、Facebook、Twitter等著名信息技术(IT)企业的

数据中心;云业务已经成为亚马逊盈利最大的部门;Dropbox的成功很大程度上也取决于亚马逊的云服务。然而近些年来,万物互联时代带来了数据传输带宽、时延,以及应用和服务性能、可靠性等方面的新需求。据估计,在不久的将来,数百亿台边缘设备将被部署,按照摩尔定律,它们的处理器速度也将呈指数级增长。通过把这些大量分布在网络边缘的空闲计算能力和存储空间整合,使其无缝为移动设备提供计算和存储支持,一种新的计算范式——移动边缘计算<sup>[2]</sup>(MEC)被提

出。MEC系统如图1所示,关键组件包括MEC服务器和移动设备(也称为终端用户、客户端、服务订购方)。MEC服务器通常是由电信运营商进行部署的、与终端用户非常接近的小型数据中心,通过网关经由互联网连接到数据中心,并且可以与无线接入点进行协同合作。移动设备和服务器可以使用先进的无线通信和网络技术建立可靠的无线链路。

MEC的出现使得传统电信蜂窝网络和互联网业务得到了深度融合,减少了移动业务交付时的端到



▲图1 MEC系统

端时延,挖掘出无线网络的潜在能力,进而提升用户体验。MEC模式将给电信运营商在运作模式上带来一系列全新的变革,继而推进电信运营商建立新型的产业链和网络生态圈。

移动边缘计算的出现同时推动了物联网、5G和运营商个性化业务的发展。

#### (1) 物联网的高速发展。

未来是万物互联的时代,国际数据公司(IDC)的统计数据显示:预计2020年将有500亿的智能终端和设备接入互联网<sup>[3-4]</sup>。连接设备的数量的快速增长,不仅代表着海量数据的产生,接入网的设备往往还需要进行一定的智能计算。物联网的核心理念是万物互联<sup>[5]</sup>,是为了让每个物体都能够智能地连接和运行。MEC遵循“业务应用在边缘,管理在云端”的计算模式<sup>[6]</sup>,完成了把计算、网络和存储能力从云延伸至物联网网络边缘的转变。通过在近网络边缘进行数据分析处理,MEC使得物联网中物与物之间

的传感、交互和控制变得更加容易。具备计算和存储能力的智能终端使得在网络边缘完成数据的分析处理成为可能。

(2) 5G网络技术需要MEC的支持。

移动互联网和物联网的快速发展推动了5G网络技术<sup>[7]</sup>的诞生,5G技术要求网络实现“大容量、大带宽、大联结、低延迟、低功耗”。在当前的网络架构中,核心网部署在远端,导致传输时延比较大,无法满足低延迟业务需求。MEC部署在网络边缘,融合了无线网络和互联网技术,并在无线网络侧增加了计算、存储等功能模块。通过把5G应用程序和业务部署在移动边缘,MEC可以减少数据传输过程中的转发和处理时间,明显降低端到端时延,并且在满足低延时要求的情况下,降低了功耗。

(3) MEC助力运营商推出个性化业务。

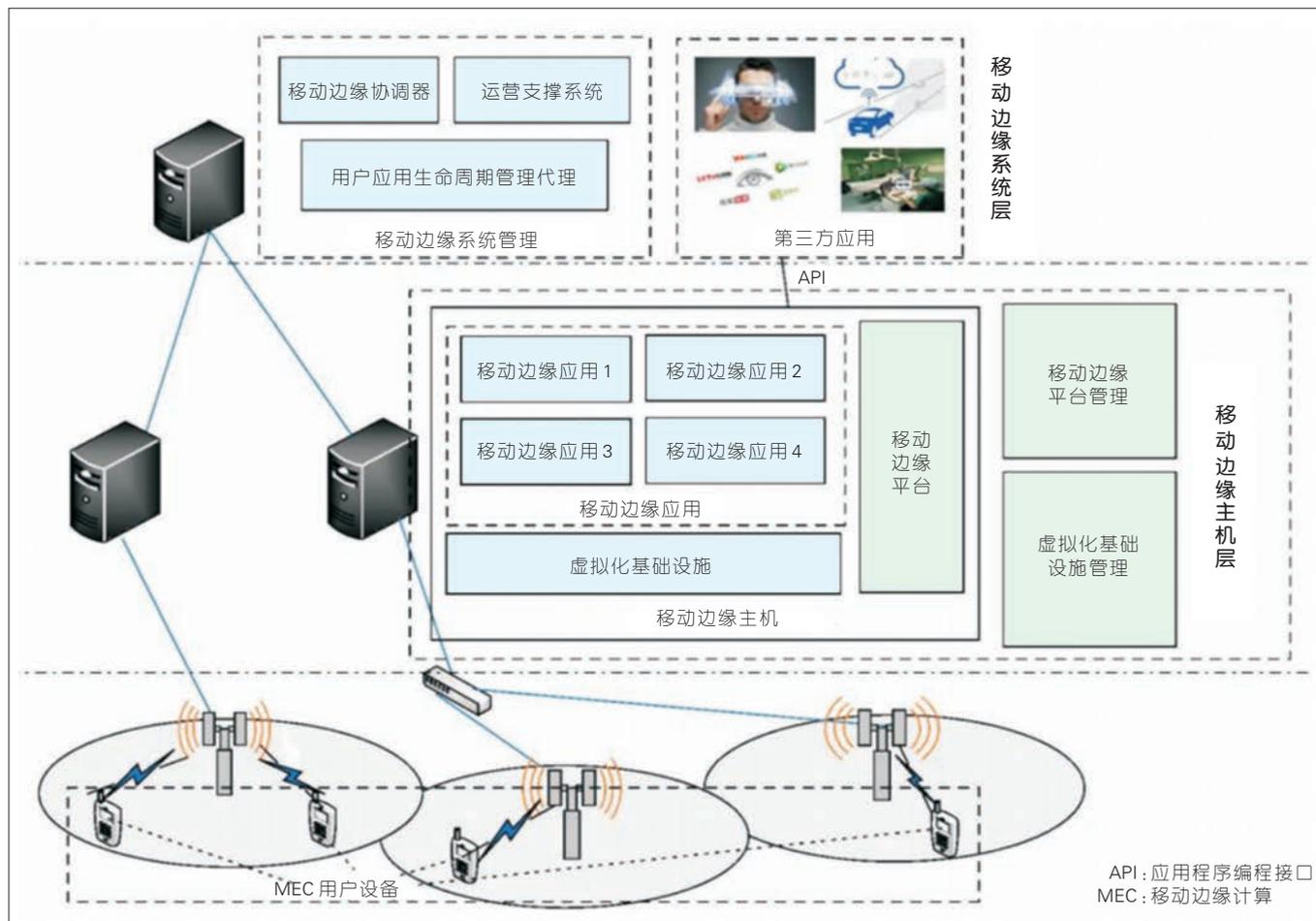
传统的运营商网络是非智能的。包月套餐大量存在,难以满足

不同用户的需求差异;资费相同的情况下,带来的结果是流量使用少的用户补贴流量使用多的用户;业务缺少优先级的区分,导致对实时性要求高的业务无法获得优先保障。与非智能的网络形成对比的是通信网络正在承载着更多基于新型智能终端、基于IP的多媒体应用,运营商商业模式和资费的单一,让运营商在业务和用户的掌控力上显得不足。MEC的关键技术包含了业务和用户的感知能力,因为边缘网络靠近用户,能够实时获取用户真实的信息。MEC可以通过大数据分析用户的特征实现用户画像。用户画像有利于对用户需求和行为进行预测,为不同的用户定制个性化业务,带来更好的用户体验。

## 2 MEC的架构

在欧洲电信标准化协会(ETSI)制定的“MEC全球标准003版本”(GS MEC 003)中,ETSI定义了移动边缘计算基于网络功能虚拟化(NFV)的参考架构<sup>[8]</sup>。根据ETSI的定义,移动边缘计算侧重的是在移动网络边缘给用户提供IT服务的环境和云计算的能力,意在靠近移动用户来减少网络操作和服务交付的时延。移动边缘计算架构分为3级:系统层、主机层和网络层,如图2所示。ETSI提出的系统架构中展示了MEC的功能要素和每个功能要素之间的参考节点。

其中,系统层结构由MEC系统级管理、用户和第三方实体组成。MEC系统级管理用于掌握部署的MEC主机、可用资源、可用MEC服



▲图2 欧洲电信标准化协会定义的 MEC 架构

务和整个网络拓扑;加载用户或第三方应用程序包,包括检查包的完整性和真实性,验证应用程序的规则和要求,必要时进行调整以满足运营商策略;记录加载的数据包并准备虚拟基础架构管理器以进一步处理应用程序,以便它可以根据应用程序处理的要求管理虚拟化基础架构,例如分配、管理和释放虚拟化基础架构的虚拟化资源;基于延迟、可用资源等选择或重新选择适当的 MEC 主机应用程序。

主机层架构主要包括 MEC 主机级管理和 MEC 主机。MEC 主机级管理包括移动边缘平台管理器

和虚拟化基础架构管理器;而 MEC 主机由移动边缘平台移动、移动边缘应用和虚拟化基础架构3部分组成。MEC 主机级管理主要进行移动边缘平台管理和虚拟化基础架构管理,移动边缘平台和移动边缘应用可以提供或使用彼此的服务,例如,移动边缘应用发现和使用移动边缘平台提供的无线网络信息、用户和其他相关的位置信息、带宽管理等服务,同时通知平台可以为用户提供服务。移动边缘平台为移动边缘应用程序提供运行环境,同时接收来自管理器、应用程序或服务的行业规则进而对数据执行

对应的指令,进行业务路由。移动边缘平台还从其管理员的域名解析系统接收记录,并配置域名系统(DNS)代理/服务器以管理移动边缘服务。例如,可以在移动边缘平台上的服务列表中注册应用程序,使其成为平台提供的移动边缘服务之一。此外,还可以通过移动边缘平台访问永久存储和时间信息。虚拟化基础设施使用通用硬件来提供底层硬件的计算、存储、网络资源和硬件虚拟化组件,以运行多个移动边缘应用程序,从而减少了处理成本,有限的资源也可以灵活有效地重复使用和共享;移动

边缘应用是基于虚拟化基础设施形成的虚拟应用程序,通过标准应用程序接口和第三方应用程序对接,并为用户提供服务。

英特尔对 MEC 的整体架构也做了定义<sup>[9]</sup>: MEC 处于无线网络接入点和有线网络之间,因为传统的无线接入网拥有业务本地化和近距离部署的优势,从而带来的是高带宽和低时延的传输能力。MEC 模式下通过将网络业务“下沉”到更加接近用户的无线网络接入侧,直接的好处就是用户能明显感受到传输时延减小,网络拥塞情况被显著控制。MEC 提供应用程序编程接口(API),对第三方开放基础的网络能力,使得第三方可以根据业务需求完成按需定制和交互,具体如图 3 所示。

### 3 MEC 的应用案例

#### (1) 增强现实(AR)<sup>[10]</sup>

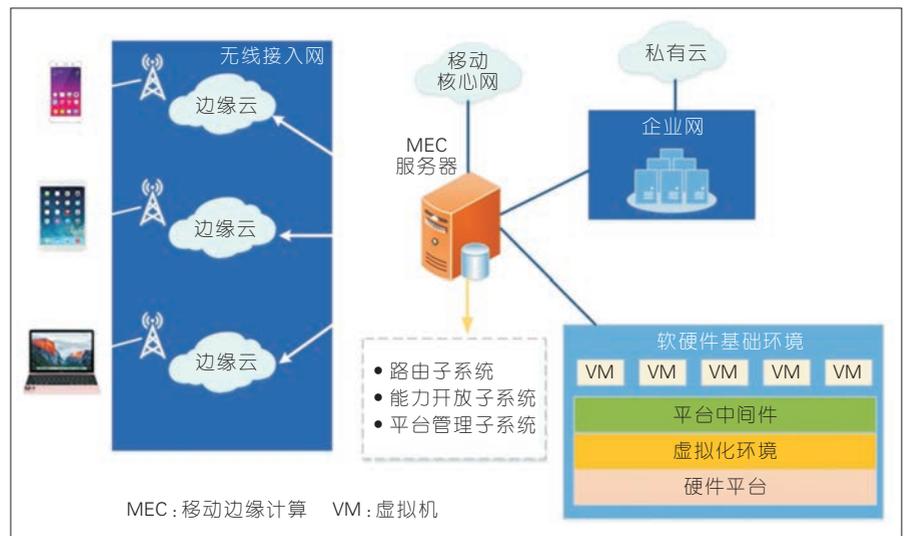
当一个支持高数据传输速率和低延迟计算的移动网络被部署的时候,新应用的出现变成了可能。AR 就是要求有着高数据传输速率和低延迟计算的应用场景。AR 技术是将真实世界环境和虚拟信息高度集成,生成被人类感官所感知的信息,

来得到超越现实的感官体验。AR 可用于智能手机、智能眼镜和平板电脑等移动设备上,以支持新的应用和新的服务,例如 3D 电影、虚拟游戏等<sup>[11]</sup>。AR 技术需要使用终端设备的相机或者定位技术,通过分析拍摄的图像来确定用户所处的位置和朝向。由于 AR 在对视频、图像等任务复杂性很高的数据进行处理的同时,还要实现和用户的实时互动,所以对时延极为敏感,对数据传输速率也有了很高的要求。MEC 将延迟敏感的 AR 任务卸载到附近的边缘计算服务器中执行,能够让应用得到更快的响应速度,同时又降低了任务的处理延迟。图 4 显示

了一个基于 MEC 的增强现实应用系统,其中 MEC 服务器能够通过准确地分析输入数据来区分请求的内容,然后将增强现实数据发送给最终用户。

#### (2) 视频优化加速。

近年来,网络速度的不断提升,带动了视频流量的飞速增长。据思科的统计,全球视频流量从 2012—2017 年,增长了接近 2.5 倍。随着 5G 商用的临近,网速将进一步得到提升,同时刺激视频流量的飞速增长。到 2021 年,移动视频将占总移动流量的 78%。移动视频流量猛增的同时,解决网络延迟的问题就变得刻不容缓。MEC 具有高带宽和



▲ 图 3 英特尔定义的 MEC 架构

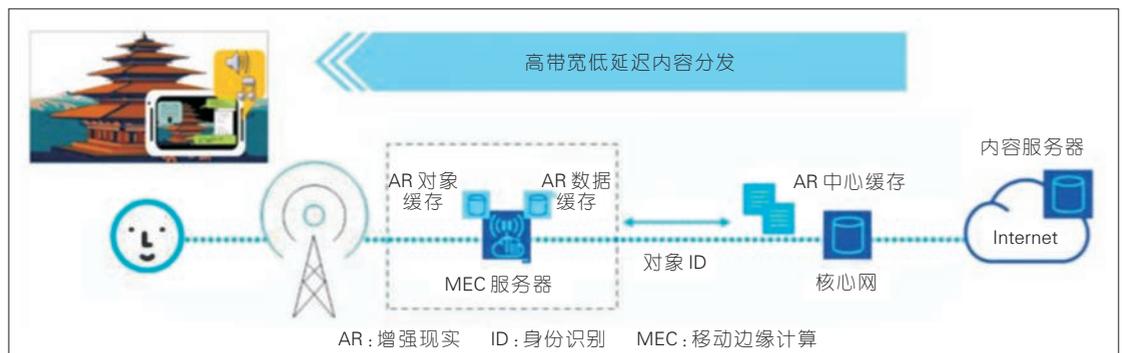


图 4 增强现实服务场景

低时延的特征,同时在本地产拥有一定的计算能力,特别适合用来对视频播放进行加速<sup>[12]</sup>,提高用户体验。如图5所示,MEC通过运用本地缓存技术,将视频内容缓存到靠近无线侧的MEC服务器上,用户一旦发起视频请求,MEC服务器先去检查本地是否有该内容,若有则直接提供服务,否则MEC服务器从Internet服务器提供商获取内容下载到本地,供其他用户访问。由于MEC服务器和Internet服务商是通过有线数据传输的,所以速度传输速率很快,减少了用户直接访问的时延。

### (3) 视频流分析。

视频流分析在车牌识别、人脸识别、家庭安全监测控制等领域有着广泛的应用,视频流分析的基本操作包括目标检测和分类。视频分析算法通常具有较高的计算复杂度,因此可以将分析工作从视频捕获设备(如摄像机)中移开以简化设备设计和降低成本。如果在中心云中去处理这些计算复杂度高的任务,视频流会被路由到核心网络中,这将消耗大量的网络带宽。通过在靠近边缘设备的地方

进行视频流分析,系统不仅具有了低延迟的优点,还可以避免大量视频流上传导致的网络拥塞问题。基于MEC的视频流分析系统如图6所示<sup>[13]</sup>,其中边缘服务器应该具有视频管理和视频分析的能力,并且只需将有价值的视频片段或者截图备份到云数据中心就可以完成。

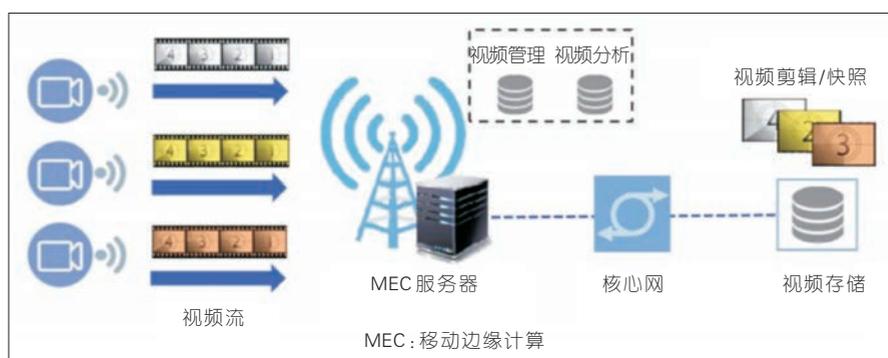
### (4) 物联网(IoT)。

为了使IoT设备的硬件复杂性降低并延长设备的电池寿命,可以把原本需要远程处理的计算密集型任务卸载到边缘服务器中去执行,并在服务器处理完后将结果返回给用户。另外,一些IoT应用程序需要获得系统的分布式信息用于执行计算任务。MEC能帮助IoT设备很好地解决这个问题。MEC服务器自身承载着高性能的计算

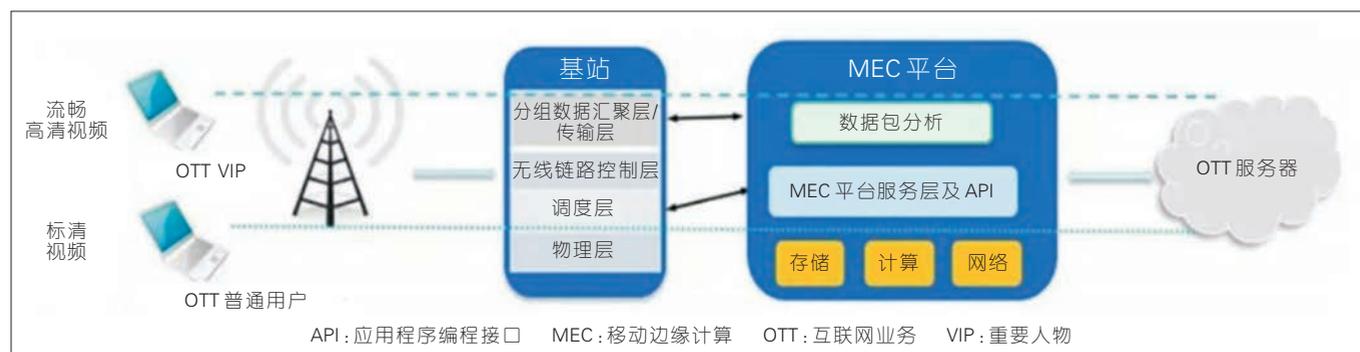
能力并能够收集分布式的信息,所以通过它们的部署能够显著简化IoT设备的设计复杂度。IoT设备可以不需要具有强大的处理能力,也不需要自己从多个数据源接收信息用来执行比较复杂的计算。IoT的另一个重要特征是运行设备的异构性,不同的终端设备运行着不同形式的协议。所有的设备应该是由一个低延迟聚合点(网关)完成管理工作,MEC服务器就可以充当这个网关的角色,如图7所示。

### (5) 车联网。

车联网技术<sup>[14]</sup>可以通过各种传感器来感知车辆的行为和路况,提高车辆的安全性,减轻交通拥堵的程度,也能带来一些增值服务的机会,比如车辆定位、寻找停车位等。这种技术还没有达到成熟,车



▲ 图6 移动边缘计算平台下的视频流处理



▲ 图5 基于MEC平台的在线视频系统

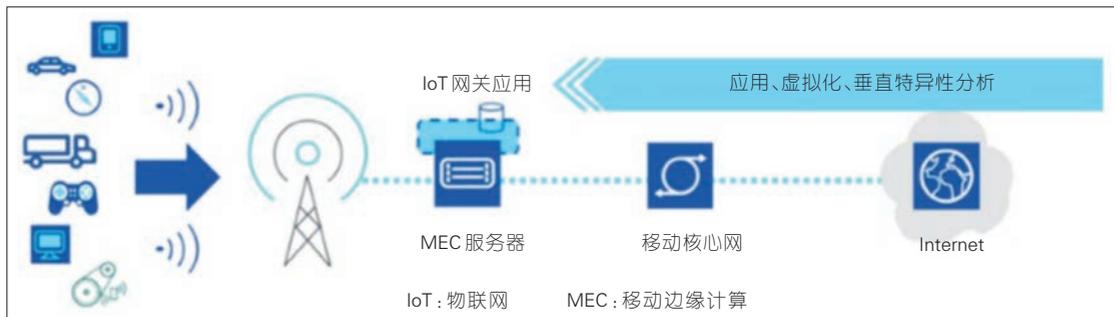


图7 物联网服务场景

到云端的延迟还处于 100 ms ~ 1 s 之间,远不能达到要求。MEC 通过将连接的车云系统扩展到高度分布的移动基站环境中,使数据和应用程序可以更靠近车辆,这样可以有效地减少数据的往返时间。MEC 在车联网中的应用如图 8 所示,应用程序运行在部署于长期演进(LTE)基站站点的 MEC 服务器上,提供道路侧的相关功能。通过接收并分析来自邻近车辆和路边传感器的消息,边缘计算能够在 20 ms 端到端的延迟内传播危险警告和延迟敏感信息。低延迟使得附近的车辆能够在几毫秒内接受到数据,从而让驾驶员可以立即做出反应。

#### 4 MEC 的挑战

MEC 作为一个新兴的前沿技术,在为我们带来便利和效益的同时,还存在很多的问题和挑战。为了促进 MEC 的发展,我们将从安全和隐私、任务资源分配和通信这三个方面重点描述 MEC 面临的问题和挑战。

##### 4.1 安全和隐私

MEC 支持新类型的服务,但是它独有的特性也带来了新的安全性和隐私问题。

##### (1) 信任和认知机制。

信任是一种几乎在每个移动系统中都存在的重要安全机制,其基本的思想是让系统认可与之交互的实体的身份。认证管理作为另一种安全的手段,也提供了一种可能的解决方案来确保“信任”。然而,MEC 系统固有的异构性,即不同的边缘服务器可以由多个供应商部署,各种移动终端设备共存,使得传统云计算系统的信任和认证机制不适用。此外,在 MEC 网络中,边缘服务器将为大量的延迟敏感型应用程序提供服务,这让 MEC 中的信任

和认证设计比传统的云计算系统复杂得多。因此,构建认证机制和设计分布式策略是亟待解决的问题。

##### (2) 网络安全。

用于支持 MEC 系统的通信技术,如 WiFi、LTE 和 5G,都有自己的安全协议来保护系统免受攻击和入侵。然而,这些协议不可避免地创建了不同的信任域。MEC 系统中网络安全的第一个挑战来自证书分发的困难性,证书是可用于在不同信任域直接进行协商的会话密钥。在现有的密钥分发方案中,证书颁发机构只能将证书分发给位于其自

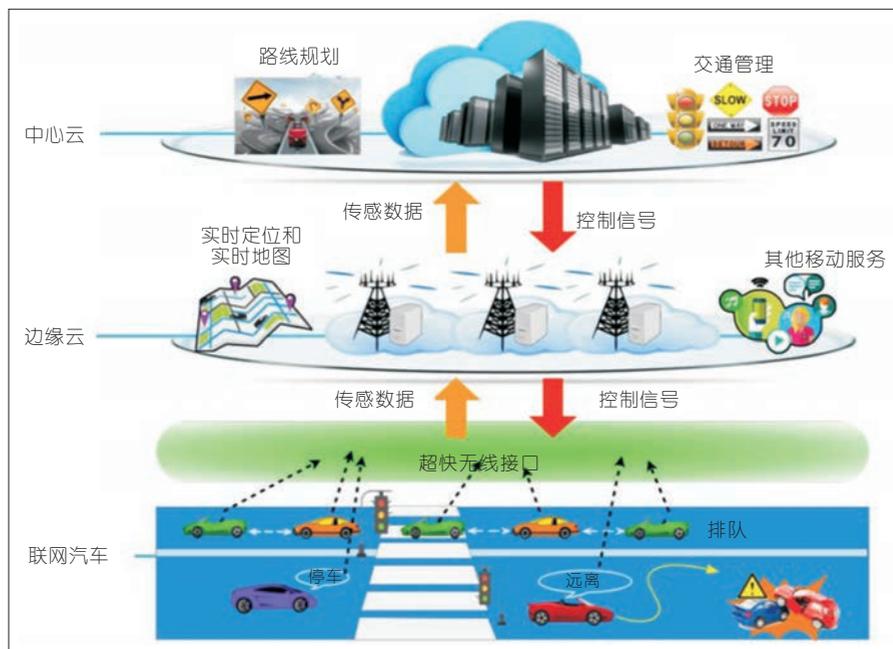


图8 移动边缘计算在车联网中的应用

身信任域内的实体,这难以保证不同信任域之间的通信隐私和数据完整性。为了解决这个问题,我们可以使用加密属性作为凭据,以便交换会话密钥。我们也可以使用联合内容网络的概念,定义多个信任域协商和维护域间凭据。此外,将软件定义网络(SDN)和NFV等技术引入 MEC 系统,以简化网络管理,并提供隔离也是一种有效的方法。但是,以上的这些技术本质上都是软件,因此很容易受到攻击。并且,MEC 系统存在着大量的设备和实体,增加了攻击者成功攻击单个设备的机会,单个设备被攻击可能造成整个系统的崩溃。所以,我们需要更加新颖和健健全安全机制,例如管理程序内省、运行时内存分析和集中式安全管理等,以保证 MEC 系统的安全联网环境。

#### (3) 安全和私有计算。

将计算密集型应用程序迁移到边缘计算服务器是构建 MEC 系统的重要功能和动机。在实践中,任务输入数据通常包含敏感的私人信息,如个人临床数据和商业财务记录。因此,在将这些数据卸载到边缘计算服务器,特别是不信任的服务器之前,应该对其进行适当的预处理,以避免信息泄露。除了信息泄露之外,边缘计算服务器还可能由于软件缺陷或经济的激励而返回不准确甚至不正确的计算结果,特别是对于具有巨大计算需求的任务。为了实现安全且私有的计算,边缘平台可以通过加密算法和可验证技术实现执行计算任务而不需要知道用户的原始数据,并且验证计

算结果的正确性。

## 4.2 任务资源分配

### (1) MEC 资源的分布与管理。

边缘计算中一个很重要的研究领域就是如何协调边缘节点的计算资源。为了保证所有希望使用 MEC 资源的用户都能够得到无处不在的服务,MEC 服务器和计算/存储资源应该分布在整个边缘网络中。因此,MEC 服务器在物理位置的放置上应该以分层的方式相互补充,这将使得计算资源和存储资源得到有效的利用,同时极大地满足用户服务质量(QoS)和体验质量(QoE)的要求。在此背景下,一个重要的挑战是根据预期的用户需求找到在物理上放置边缘计算服务器的最佳方式,同时考虑运营成本。

我们还应该为 MEC 资源的合理管理设计高效的控制程序,这包括信令消息的设计、信令的交换以及信令开销方面的优化。控制信息应该能够传送状态信息,例如单个计算节点的负载和无线/回程链路的质量,以便有效地编排 MEC 内的计算资源。如果传送的状态信息很少被交换,那么我们应该在与状态信息频繁交换有关的高信令开销和由于状态信息老化而对 MEC 性能的影响之间找到平衡点,并且提出有效的信令机制,以确保 MEC 中的控制实体信息能即时更新,同时使得获得这些信息的成本最小化。

### (2) 计算资源的分配。

针对卸载到 MEC 的应用程序的计算资源如何有效分配的问题,目前的研究没有考虑网络的动态

性。更确切地说,现在的分配策略是将应用程序卸载到 MEC 之前预先选择计算节点,然后假定相同的计算节点处理卸载的应用程序。然而,MEC 在处理完给定应用程序时会释放一些额外的计算资源,这些资源是还可以给其他任务提供服务的,以便进一步加速卸载的计算过程。因此,在 MEC 中处理卸载的应用程序期间动态地分配计算资源是未来需要解决的一个研究挑战。

## 4.3 通信

引入的 MEC 服务器对第 3 代合作伙伴计划(3GPP)<sup>[15]</sup>网络体系结构和现有的接口中应该具有透明性。用户设备(UE)和核心网络元件在符合现有 3GPP 规范前提下应该不受 MEC 服务器和托管在服务器上的应用程序的影响。3GPP 协议和程序应该在不影响服务协议(SLA)的情况下操作和运行。

### (1) 应用可移植性。

在 MEC 中,对应用程序的一个基本要求就是程序能被无缝地在不同供应商提供的 MEC 平台中加载和执行。应用的可移植性消除了平台基础上进行复杂的开发和集成工作,减少了软件应用程序开发人员的很多麻烦。移动边缘计算允许在 MEC 服务器之间快速地进行服务迁移,从而优化了服务器资源。应用平台提供的可弹性扩展服务是确保应用程序可移植性的关键。平台管理的框架需要在不同的解决方案之间保持一致,来确保即使在不同的管理环境下应用程序开发者也能简单地进行移动边缘计算方面的

工作。用于部署和管理应用程序的工具和机制也需要在各个平台和供应商之间保持一致,一致性使得软件应用开发人员能确保他们的应用程序管理框架无缝集成。

(2)移动性。

移动性问题又可以分为以下2种场景:

- 用户在一个 MEC 服务器之下,从一个基站移动到另一个基站。服务器需要保证用户到应用的连接性能,需要跟踪用户当前的连接节点来确保下行数据的路由。

- 用户从一个 MEC 服务器移动到另一个 MEC 服务器。这个过程需要基于应用的能力、原始服务器和目标服务器的负载信息等情况来决策采用的移动性机制。

到目前为止,对移动性的研究侧重于虚拟机(VM)迁移中的移动性管理,主要考虑的是只有一个计算节点为每个用户设备提供计算服务。因此,当应用程序卸载到多个计算节点时,如何有效地处理 VM 迁移过程将成为一个挑战。此外,VM 迁移给回程链路带来高负载,并导致了高延迟的发生,这使得 VM 迁移不适合实时应用。因此,开发新的高级技术,实现以毫米级别的速度进行 VM 的快速迁移是将来的一个研究方向。例如,可以基于一些预测技术达到预先迁移计算,以使用户不会察觉到服务中断。同时,也可以更仔细地研究动态优化和联合考虑所有用于移动性管理的技术(例如功率控制、VM 迁移、迁移数据的压缩和路径选择),以便提高用户设备的 QoE 并

优化总体系统性能。

## 5 结束语

MEC 作为一种新的计算范式,一方面可以满足用户设备飞速增长的计算需求,另一方面也可以提高用户体验的质量。MEC 通过将大量的计算资源和存储资源网络边缘化,给用户终端提供类云计算的功能和 IT 服务。移动设备和边缘服务器之间通过无线通信直接交互,可以支持具有超低延迟要求的应用,延长设备的电池寿命,让网络运营变得高效。

## 致谢

本研究得到工业和信息化部 2018 年工业互联网创新发展工程“浪潮云工业互联网平台试验测试项目”的支持。在文章撰写过程中得到北京邮电大学王茜博士的帮助,谨致谢意!

## 参考文献

- [1] MIETZNER R, LEYMAN F. Towards Provisioning the Cloud: On the Usage of Multi-Granularity Flows and Services to Realize a Unified Provisioning Infrastructure for SaaS Applications[C]//2008 IEEE Congress on Services-Part I. IEEE: USA, 2008: 3-10. DOI:10.1109/SERVICES-1.2008.36
- [2] 施巍松,孙辉,曹杰,等.边缘计算:万物互联时代新型计算模型[J].计算机研究与发展,2017,54(5): 907-924. DOI:10.7544/issn1000-1239.2017.20160941
- [3] 5G 愿景与需求(白皮书) [EB/OL]. (2016-04-26) [2019-12-28]. <http://www.docin.com/p-890406206.html>
- [4] NGMN 5G White Paper[EB/OL]. (2016-04-26) [2019-12-28]. <http://www.ngmn.org/5g-white-paper.html>
- [5] ASHTON K. That 'Internet of Things' Thing [J]. RFID Journal, 2009, 22(7): 97-114
- [6] 王宏宇.边缘计算在智慧城市中的应用[J].电脑迷,2018,(8): 138
- [7] Mobile Edge Computing-A Key Technology Towards 5G[EB/OL]. (2016-04-26)[2019-12-28]. <http://www.etsi.org/technologiesclusters/technologies/mobile-edge-computing>
- [8] 罗雨佳,欧亮,唐宏.基于 NFV 的边缘计算承载思路[J].电信科学,2018,34(8): 153-159

- [9] 俞一帆,任春明,阮磊峰,等.移动边缘计算技术发展浅析[J].2016,11(11): 59-62
- [10] PANKRATZ F, KLINKER G. [POSTER] AR4AR: Using Augmented Reality for Guidance in Augmented Reality Systems Setup[C]//2015 IEEE International Symposium on Mixed and Augmented Reality. USA: IEEE, 2015: 140-143. DOI:10.1109/ISMAR.2015.41
- [11] JAIN P, MANWEILER J, CHOUDHURY R R. OverLay: Practical Mobile Augmented Reality [C]//Proceedings of the 2015 International Conference on Mobile Systems, Applications, and Services. USA: ACM, 2015:331-344
- [12] PATEL M, NAUGHTON B, CHAN C, et al. Mobile-Edge Computing Introductory Technical White Paper[J]. White paper, Mobile-Edge Computing (MEC) Industry Initiative, 2014: 1089-7801
- [13] European Telecommunications Standards Institute (ETSI). Mobile-Edge Computing Introductory Technical White Paper [EB/OL]. (2016-12-03)[2019-12-28]. [https://portal.etsi.org/Portals/0/TBpages/MEC/Doc/Mobile-edge\\_computing\\_-\\_Introductory\\_Technical\\_White\\_Paper\\_V1%2018-09-14.pdf](https://portal.etsi.org/Portals/0/TBpages/MEC/Doc/Mobile-edge_computing_-_Introductory_Technical_White_Paper_V1%2018-09-14.pdf)
- [14] GUBBI J, BUYYA R. Internet of Things (IoT): A Vision, Architectural Elements, and Future Direction [J]. Future Generation Computer Systems, 2013, 29(7): 1645-1660. DOI: 10.1016/j.future.2013.01.010
- [15] 3GPP. Study on Facilitating Machine to Machine: TR 22.868 V8.0.0. 2007.03[S]

## 作者简介



**高志鹏**,北京邮电大学教授、博士生导师,同时担任国家自然科学基金委、科技部、工信部项目评审专家;目前研究方向为边缘计算、群智计算与大数据、区块链应用技术、云平台管理、网络管理与通信软件等;研究成果获省部级科技一等奖3次;已发表论文50余篇,拥有20余个国家发明专利、4项国际标准和多项行业/企业标准。



**尧聪聪**,北京邮电大学在读硕士研究生;研究方向为边缘计算系统资源分配机制。



**肖楷乐**,北京邮电大学在读博士研究生;研究方向为边缘计算中的任务卸载和资源分配问题、区块链应用等。



# 基于边缘计算的物端系统挑战与愿景

## Challenges and Vision of Things System in Edge Computing

彭晓晖/PENG Xiaohui, 徐志伟/XU Zhiwei

(中国科学院计算技术研究所, 北京 100190)

(Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China)

**摘要:** 针对物端计算系统面临的“昆虫纲悖论”和“数据过于集中”2大生态问题,从计算系统的角度,讨论分析了其在系统架构、应用层协议、资源管理、开发与运行时、计算负载调度5个方面面临的挑战。提出了物端控制域计算机设计愿景,即将物端计算系统划分为具有相同或兼容管理策略的控制域,并将每个域看作一台分布式计算机,为其设计物端信息总线架构原型。

**关键词:** 边缘计算;物端计算;物端控制域计算机;体系架构;应用层协议

**Abstract:** To address the two ecosystem problems of things system: “Paradox of Classis Insecta” and “Centralized Data”, five challenges of things system in perspective of computing system are discussed in this paper. These challenges include system architecture, application protocol, resource management, application development and runtime environment, and computing payload scheduling. Then, the design vision of the physical-zone-of-control computer is put forward, which divides things system into multiple zones of control with same or compatible management policy. The physical information bus architecture prototype which regards every zone as the distributed computer is finally proposed.

**Key words:** edge computing; things computing; physical zone of control computer; system architecture; application protocol

DOI: 10.12142/ZTETJ.201903005  
网络地址: <http://kns.cnki.net/kcms/detail/34.1228.TN.20190529.1507.004.html>

网络出版日期: 2019-05-29  
收稿日期: 2018-12-22

## 1 物端系统的背景

### 1.1 万物互联与物端计算系统

信息技术正在进入人机物三元计算的万物互联新时代,计算设备正在被赋予了各种具体的形态并融入到物理环境中<sup>[1]</sup>,我们称这样的设备为物端设备,例如温湿度传感器、摄像头等。物端设备正以指数级增长的方式连入互联网,形成物端计算系统,是边缘计算<sup>[2]</sup>

的重要组成部分。物端系统不断地从物理世界感知多源异构的数据,传输到信息世界进行计算,信息世界数据处理的结果也通过其反馈到物理世界,物端系统是感知和控制数据的出入口,是未来人机物三元计算空间的咽喉要塞。物端计算系统天生位于网络的边缘,具有天然的近数据源优势。将计算迁移至物端系统可以大幅减少网络带宽消耗,缩短应用响应延迟和保护用户

数据安全和隐私。

物端计算机的显著特征是需要与物理世界进行交互,不断地感知物理世界状态,并通过反馈控制改变物理世界状态。例如,空气净化器可以根据小于 $2.5\ \mu\text{m}$ 的细颗粒物(PM 2.5)传感器数据,调整净化器的净化强度,直至达到设定的目标。这是人机物三元空间中典型的感知-计算-控制循环,也是未来万物互联时代的基本计算模式之一。

其中,控制环节可以是系统根据某种策略或算法自动执行,或由用户根据计算结果选择控制动作。

在现有的物联网架构中,由于物端计算机的软硬件平台还未完全准备好计算负载迁移至离环境和人更近的物端上,该循环的计算环节通常在云端进行,造成应用的响应延迟和不确定性。人机物三元计算环境中,大量的应用对实时性有较为苛刻的要求。我们在利用亚马逊的Alex语音控制打开卧室照明灯的实验中,每次开灯平均耗时3 s以上<sup>[9]</sup>。如图1所示,智能音箱采集的语音数据传输到位于美国的亚马逊云计算服务中心进行解析,识别出的开灯命令发送到位于新加坡的设备控制服务器,通过应用程序编程接口(API)打开位于本地的照明设备。由于数据通路过长,云计算模式显著增加了开灯操作的时延和不确定性。

## 1.2 物端计算系统的生态问题

物端计算系统面临的第一个生态问题是数据过于集中。随着联网

的设备不断增加,其产生的数据也是呈指数级增长。据国际数据公司(IDC)预测,到2025年,全球数据年产量将达160 ZB,其中75%以上是产生于网络边缘和物端设备<sup>[4]</sup>。近年来人工智能和大数据的发展显示了用户数据在下一代信息系统中的重要作用,以至于被誉为信息时代的“新石油”。然而各大设备厂商均将物端设备感知到的用户数据收集到私有或公有的云计算中心存储和分析。边缘计算、雾计算和物端计算等新的计算模式均旨在将云计算中心的负载向下迁移,这种迁移目的之一也是为了解决数据过于集中的问题。万维网(Web)发明者Tim Breners-Lee在2018年提出了去中心化平台Solid,以解决数据过于集中带来的一系列问题。

物端计算系统面临的另外一大生态问题是昆虫纲悖论<sup>[1]</sup>,即物端设备的种类像昆虫一样繁多,其本质描述的是系统的多样性和无序性。多样性体现在应用场景、芯片架构、硬件接口、操作系统、应用软件、编程模式、通信协议等多个方

面。这造成了物端计算机应用开发、部署和运维十分困难,程序员通常需要熟练掌握从硬件选型、软件开发到通信协议选择的全栈知识,才能研发出高效、可靠的新产品。无序性表现在系统的架构、设计方法、编程模式等方面。由于缺乏统一的系统架构、设计和编程方法,物端应用变得十分碎片化,设备之间难以协同计算,从而对人机物三元空间人的行为做出适当的反应。

物端设备面向场景设计、功能较为单一、计算资源十分有限,是一台专用计算机。因此,并不需要太多的计算、存储和通信资源。随着人类生活水平的提高,对物端设备的能力要求也逐渐提高。例如,一些空调已经具备了无线联网的功能,而冰箱甚至有对内存食物分析,并结合家庭用户的健康数据推荐合适的采购清单的功能。因此,我们需要一个统一的、开放的系统技术栈和统一的、协调的分布式系统体系结构抽象来支持物端计算机和其分布式系统的具体设计与实现,从而承担由云端迁移下来的计算负载。

## 2 物端计算系统面临的问题

### 2.1 物端系统体系结构

本文所述体系架构包括2个方面:单个物端设备的软硬件系统结构和多个物端设备组成的物端域分布式架构。个人电脑(PC)Web和移动Web已经证明:应用的繁荣可以直接推动相关领域的快速发展。智能手机应用(APP)和应用商



▲图1 亚马逊智能音箱开灯命令执行过程<sup>[9]</sup>

店(APP Store)的出现,使在手机上开发和部署应用的门槛大大降低,开发效率大大提升,促进了功能手机到智能手机的转变。物端计算系统的多样性和无序性已经严重影响了其应用开发和部署的效率,需要有合理的、统一的系统结构与编程方法,来支撑应用的高效开发和部署。而且,物端计算系统多是具有一定物理或概念边界的小型分布式计算系统,其开发环境需要分布式技术作为支撑,这进一步加大了系统的设计与应用开发复杂度。

自Web发明以来,信息系统经历了PC互联网时代、移动互联网时代,即将进入万物互联网时代。PC互联网时代经历过数十年的发展终于形成了x86+Windows/Linux的单机系统架构和以表述性状态传递(REST)原理为基础的分布式系统架构<sup>[3]</sup>。这种统一架构促进了PC应用的爆发,从而使计算机应用到了各行各业,提高了人类的生产效率和生活水平。经过10多年的发展,智能手机也形成了ARM+Android/iOS的单机系统架构和以REST为基础的分布式系统架构<sup>[4]</sup>。当前智能手机上很多应用代码均运行在服务器端,而手机端只需一个“浏览器”来向用户呈现服务器端计算的内容。20年来Web系统的成功证明:由超文本传输协议(HTTP)设计和开发实践总结而来的REST分布式架构基本原理<sup>[5]</sup>,是解决分布式系统多样性和可扩展性的最佳实践之一。因此,将REST原理拓展至万物互联的物端计算系统是一个潜在的解决方案。

然而,物端的多样性和无序性远远高于PC互联网和移动互联网系统,目前并无统一开放的单体系统架构和分布式架构来支撑物端运行Web的协议栈。物端有普通的微控制单元(MCU)、低功耗ARM、数字信号处理(DSP)处理器、神经网络加速器等多种芯片,以及集成电路总线(I2C)、串行外设接口(SPI)、控制器局域网络(CAN)总线、串行接口(UART)等多种外设总线接口。面向不同的应用场景和硬件资源限制,其操作系统更是多达几十种。它们多采用无线的方式连入互联网,根据连接距离、速度、功耗等具体要求,可选择的通信协议也多达十数种。因此,程序员通常需要熟练掌握软硬件技术栈相关知识,才能快速开发出高效、稳定的物端应用程序。学术界为了让物端计算系统更好承接计算任务的迁移,在架构和系统设计层面做了出色的工作,例如EdgeOS\_H<sup>[6]</sup>、BOSS<sup>[7]</sup>等。文献[3]中,作者提出了 $\Phi$ -Stack,通过软硬件高度协同设计的方法为物端设备提供开放式技术栈,并支持Web和智能应用。文献[8]中,作者提出了物端分布式系统架构风格,通过拓展现代Web系统的基本架构约束REST,提出物端表述状态转移(T-REST)架构原理。

## 2.2 基于传输控制协议(TCP)/用户数据协议(UDP)的物端应用层协议、中间件

应用层协议是分布式系统的重要组成部分,例如HTTP协议族,是整个互联网系统架构的核心技术之

一。由于HTTP协议族开销过大,海量资源受限的物端设备通常无法运行该协议栈。而且,许多物端系统的应用对实时性要求较高,HTTP协议通常无法提供实时性保障。因此,为物端系统设计一套可伸缩、可扩展的物端设备应用层协议是一项十分具有挑战的工作。由于现存大量的基于HTTP协议族的网络基础设施,新协议的设计需兼容Web系统的REST架构约束。

为此,学术界和产业界做了大量的工作,这些工作多数基于TCP/UDP来设计。例如,面向资源受限环境和物联网设计的受限应用协议(CoAP)<sup>[9]</sup>和消息队列遥测传输协议(MQTT)<sup>[10]</sup>。“统一接口”是REST架构风格的重要设计原则之一,HTTP的GET、PUT、POST、DELETE等请求方法是“统一接口”设计原则的重要体现。CoAP遵循REST架构约束,并继承了HTTP协议的GET、PUT、POST和DELETE 4个主要请求方法,因此与现代Web系统有很好的兼容性。为了减少开销,CoAP网络层采用的是UDP协议,通常与基于IPv6的低速无线个域网标准(6LoWPAN)<sup>[11]</sup>一起构成资源受限设备的网络通信协议栈。MQTT是为物联网设计的轻量级、支持发布/订阅模式的应用层协议,其构建于TCP/IP协议之上,可在一定程度上与现有的互联网系统兼容。MQTT的发布/订阅模式可有效支持一些物端应用场景的实时性要求,支持应用协议层的双向通信。然而,不足之处是MQTT并不完全遵循REST架构约束,用其构建的应用与

Web应用兼容性较差。

另外,学术界也提出了一些基于REST风格的协议或中间件,例如UbiREST、TinyREST、SeaHTTP等,这些研究目前尚未大规模应用。UbiREST是遵循REST架构风格,为泛在计算网络设计的、面向服务的中间件,提供服务在运行时动态组合等高级功能。TinyREST是为智能家居的消费电子和传感器组成的网络设计的应用层协议,该协议用于Web系统的HTTP请求和TinyOS的消息之间的转换。SeaHTTP则通过实现BRANCH和COMBINE2个新方法,实现请求的拆分和合并,从而提高Web应用的执行效率。

MQTT和CoAP是物端设备的应用层协议的2个代表,已经在产业界和研究机构应用。基于发布-订阅模式的MQTT对实时性和服务器/客户端双向通信有着较好的支持,而遵循面向资源的REST架构风格所设计的CoAP协议则与现有的Web系统应用有着比较好的兼容性。

### 2.3 资源管理与发现

据IDC预测,到2025年,每个人每天与周围的设备交互4800次,平均每18s交互1次<sup>[4]</sup>。这些服务进行管理和及时发现将是一个巨大的挑战。在万物互联时代,人们总是在不同的计算场景中切换,每一个场景即是一个具体的物端计算系统,彼此独立又相互联系。例如,在出差场景中,人们可能要连续经过智能家居、智慧交通、智能宾馆、智

能办公室等多个场景切换,每个场景都有一定的边界和自己的管理控制策略。人在这些场景中移动时,如何快速地对相关分布式资源进行注册、发现和销毁是提供无缝智能服务的关键。

现代Web系统的资源命名机制主要基于统一资源标识符(URI)标准,然而其并不具有严格的唯一性和可靠性。一个基于URI的URL链接所指向的资源有可能并不唯一,且很容易在互联网的交互访问过程中被第三方劫持和替换,从而指向劫持者设定的资源,形成安全隐患。除了传统Web系统的资源以外,物端计算系统资源还包括物理世界的实体,其通常需要和物理世界交互,一旦出现问题,极易威胁人们的生命财产安全;因此,需要更加严格的资源命名与管理策略,确保名称对资源的唯一标识和安全的访问。对物体的命名与标识是传统物联网的基础研究之一,代表成果有国际标准化组织(ISO)/国际电工委员会(IEC)18000、美国的EPCglobal和日本的Ubiquitous ID。

### 2.4 开发与运行时环境

应用是推动计算技术领域发展的核心动力之一。为了提高应用的开发和执行效率,开发与运行时环境是核心支撑技术。具体包括应用程序语言设计、开发与调试环境、运行时环境等。物端计算系统的多样性决定了其应用开发的困难,程序员经常要面对数十种硬件、系统软件、通信协议选择,并与具体场景适配。同时,物端计算系统通常是场

景驱动的小型分布式系统,计算负载通常需要在系统内部分割、调度和迁移,即对分布式应用的支持。

在程序语言方面,物端计算系统需要一种新的语言,在语言层面提供对物理世界感知和控制的功能。而且,人机物三元计算空间中,人(即用户)的角色十分重要,“感知-计算-控制”的计算模式通常需要人的干预。普通用户能在人机物三元空间高效地开发部署物端应用十分重要。物端计算系统的程序语言应具有比C/Java等语言更好的易用性。为了支持其分布式执行,在语言层面上还应支持程序的可分割性,即编写的应用程序可在编译或解释执行阶段被分割为更小的执行单元。

在开发环境方面,需要设计实现具有很少编程经验的程序员,甚至普通用户都可以高效编写应用的开发环境。物端计算系统的设备高度异构和多样化,传感器和控制器通常经过多种接口与物端设备相连。如何为开发环境抽象、设计一个中间层,向下屏蔽这些接口等资源,向上为应用开发提供统一的API接口,是物端应用开发与部署面临的重大挑战。物端设备是资源受限的,且处在动态的环境中,应用需要使用环境中的实时数据,这给应用的调试带来挑战。应用调试通常需要中断运行,这时从物理世界感知的数据有可能已经失效,当程序继续运行时可能无法得出正确的结果。

在运行时方面,需要设计高效的解析、执行环境,以适应大部

分物端设备资源受限的状况。为了提高应用的可移植性,物端计算系统软件架构通常分成多个层次,包括硬件抽象层、驱动、操作系统、应用支撑层、应用程序等。每一层都会带来功能冗余和额外开销,大大降低了应用的执行效率。物端计算系统设备计算能力各异,智能家居机器人、自动驾驶汽车的计算能力可媲美当前主流的计算机,而空调、冰箱等设备计算能力有限。因此,我们需要设计可灵活伸缩和扩展的模块化运行时环境,支持拥有不同计算能力的设备,在物端计算系统内部形成统一的运行时环境,从而更好地支持计算负载在不同设备上的调度。

## 2.5 计算负载的调度挑战

面向异构环境的分布式调度相关研究在云计算领域已经大量存在,是网格、云计算里基本的研究课题,而基于有向无环图(DAG)模型的分布式调度更是云计算的研究热点。物端计算系统是局部化的分布式系统,同样存在计算任务调度的问题,其调度与云端有明显不同。

首先,调度的目的不同。资源利用率最大化、吞吐量最大化、总费用最低、费用公平是目前云计算中多DAG任务调度的主要优化目标。物端计算系统是一个有一定“边界”的小型分布式系统,其计算以用户为中心,用户体验差的应用是没有价值的。因此,其任务调度的核心目标是减小应用响应延迟,提高用户体验,同时兼顾降低计算和传输的功耗。

其次,云计算中的任务划分多为“均分”的形式,其划分后形成的DAG任务图,子任务之间虽有依赖关系,但任务本身的异构性不大(对资源的需求差别不大);因此,无须对划分后的子任务进行详细的评估,计算其需要的具体资源。而在物端,其任务图通常十分异构。如何对任务进行合理的划分,对任务计算量进行精细评估也是物端计算系统任务调度的关键挑战之一。

第三,物端计算系统更加异构,资源建模更加复杂。云计算中心多是一些大型数据中心,其软硬件异构程度并不高。例如,当前机器学习采用的GPU服务器集群,通常有数十块甚至数百块型号相同的GPU和通用处理器中央处理器(CPU)组成,其异构程度并不太大。服务器之间多数通过光纤连接,通信速度对计算和调度的影响也较小。物端计算系统硬件的计算、存储和通信能力差异十分巨大。如何建立高效的分布式资源管理和评估模型,为计算负载的调度提供依据是物端计算系统的一项十分具有挑战的任务。

## 3 物端控域计算机

### 3.1 物端控域计算机架构

物端计算系统是由面向物理世界、具有物理功能的终端和通信设备及其环境组成的场景驱动的、具有一定边界的分布式系统。其显著特征是系统的多样性和设备的资源相对受限。可通过控域理论将其划分为具有类似或相同管理策略的

域,我们也称这样的域为物端控域计算机。

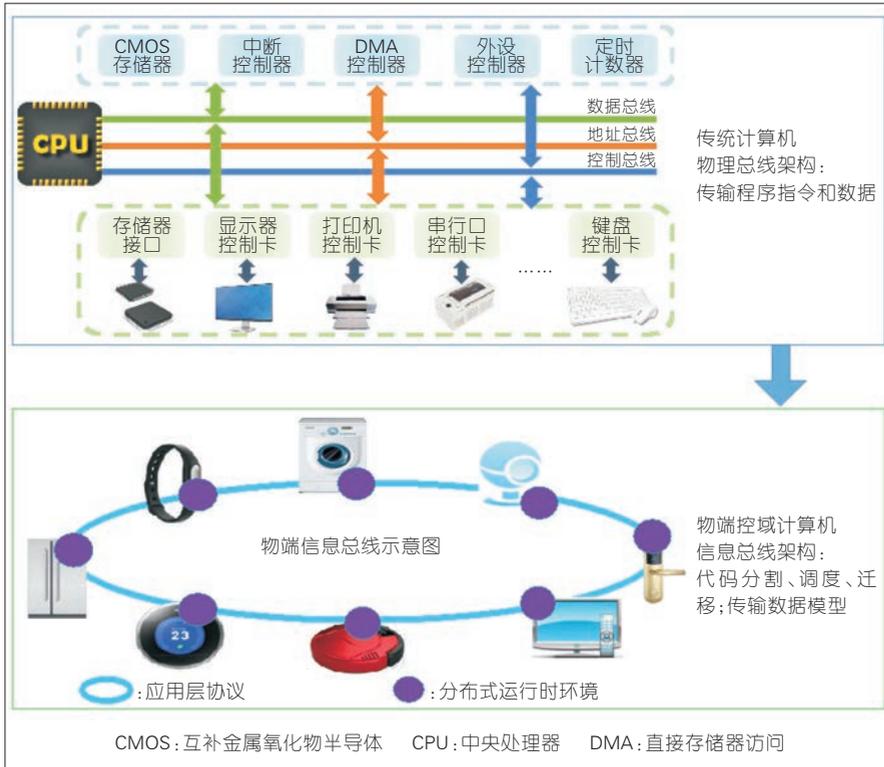
物端控域内众多异构的设备可看作该计算机的部件,它们在物理上是分散在环境中的,通过有线或无线的协议连接成“一台分布式计算机”。物端控域计算机是万物互联的不同场景抽象出来的概念。通常在这个“计算机”内有着统一或相似的控制和安全策略。我们倡导继承现代Web系统的基本架构原理REST,并针对物端特性做加强和拓展,设计适合物端控域计算机的新架构——物端信息总线架构,从而尝试解决这些问题。

### 3.2 物端信息总线

如图2所示,传统计算机物理上由不同的部件和外设通过总线连接而成,总线上传输的是程序指令和数据。而物端控域计算机的物端信息总线上传输的是微任务和数据模型,其中微任务是其计算负载的表现模式,数据模型是比原始数据更高一层的抽象。物端控域计算机的硬件由域内设备硬件和物理层通信协议组成,软件由各设备的系统软件、分布式应用运行时环境和分布式应用层协议组成。总体设计思路是将域内资源池化,统一域内软硬件资源管理和访问控制模式,并支持计算负载的分割、调度和在域内设备间自由迁移。

### 3.3 计算负载调度

物端设备远不如云端的计算资源丰富,计算负载从云端迁移到物端后通常无法由单个设备承担;因

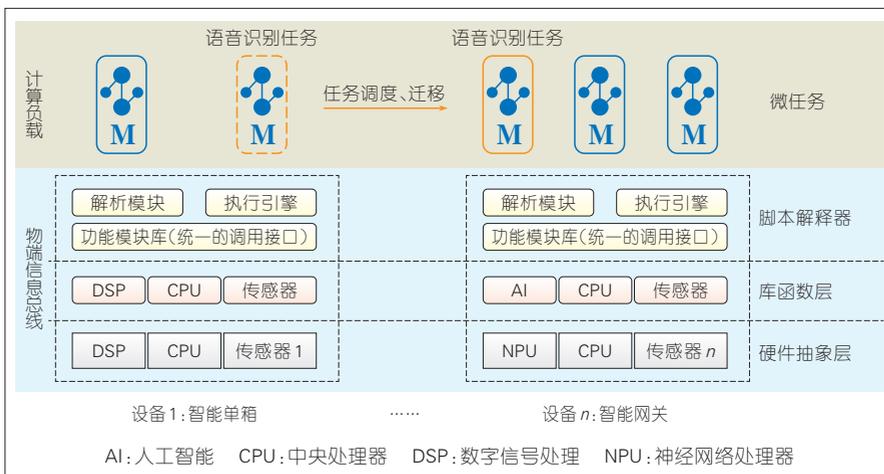


▲图2 物端控域计算机信息总线与传统计算机物理总线对比示意图

此,计算负载应具有运行时动态分割、域内自由迁徙和调度的能力。如图3所示,在智能家居的语音交互场景中,智能网关拥有对语音处理加速的硬件设施。语音需要先经过智能音箱的DSP采集和处理,得到的语音数据传输给智能网关进行

解析并得出控制命令,然后反馈给具体需要控制的设备执行。在这个过程中,智能音箱由于没有语音处理加速器,不适合处理语音解析的任务,需将处理任务迁移到具有语音处理加速器的智能网关处理。

这种计算负载的调度,除了需



▲图3 物端控域计算机及其负载调度示意图

要语言和支持外,还需要调度算法的支持。物端计算系统应用大多对响应时延较为敏感,而且对系统能耗要求更加苛刻。本文第2章已经详细阐述了物端计算系统的调度难题,调度的首要目标是满足不同计算负载的时延要求同时,最大限度降低系统的总功耗水平。而且,需要预留一部分计算资源,从而尽快处理紧急任务。

#### 4 结束语

本文阐述了物端计算系统的概念,并讨论了该领域亟需解决的5个关键问题。物端设备位于网络边缘的末端,是处理人机物三元空间计算负载的理想场所。在即将到来的万物互联时代,物端计算系统将无处不在,人们的生活、学习、工作、甚至娱乐活动均在这些不同的物端计算场景之间移动。这些计算环境是高度动态的,场景之间的资源和访问控制策略各不相同,而且物端计算系统面临多样性和无序性的现状。同时,大部分设备资源受限,需要一套新的基础架构理论和系统设计方法来具体指导其软硬件的设计和实现。

本文基于控域理论和具体挑战问题分析,提出了物端控域计算机的愿景。通过研究物端控域计算机的分布式架构、物端信息总线(即分布式协议和运行时环境)、程序语言和计算负载调度策略,以域内资源池化为主要设计思想,构建物端控域分布式计算系统,统一调度和管理域内资源。物端控域计算机将在

➡下转第57页

# 万物互联背景下的边缘计算安全需求与挑战

## Security Requirements and Challenges in Edge Computing for Internet of Everything

马立川/MA Lichuan<sup>1,2</sup>, 裴庆祺/PEI Qingqi<sup>1,2</sup>, 肖慧子/XIAO Huizi<sup>1,2</sup>

(1. 西安电子科技大学综合业务网理论及关键技术国家重点实验室, 陕西 西安 710071;

2. 西安市移动边缘计算及安全重点实验室, 陕西 西安 710071)

(1. State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an 710071, China;

2. Xi'an Key Laboratory of Mobile Edge Computing and Security, Xi'an 710071, China)



**摘要:** 作为万物互联时代新型的计算模型,边缘计算具有的分布式、“数据第一入口”、计算和存储资源相对有限等特性,使其除了面临信息系统普遍存在网络攻击之外,还不可避免地引入了一些新的安全威胁。为此,从4个方面对边缘计算的安全需求进行阐述,同时对其主要安全技术的设计以及实现所面临的挑战进行分析,较为全面地指出了边缘计算在身份认证、访问控制、入侵检测、隐私保护、密钥管理中存在的具体安全问题。

**关键词:** 万物互联;边缘计算;安全需求;隐私保护

**Abstract:** As a new computing paradigm in the Internet of Everything, edge computing has the following characteristics: distributed, data first entry, relatively limited of computing and storage resources. These characteristics bring some new security threats and network attacks to the edge computing. In this paper, the security requirements of edge computing are elaborated from four aspects, and the design of its main security technology and the challenges are also analyzed. It comprehensively points out the specific security issues of edge computing in identity authentication, access control, intrusion detection, privacy protection and key management.

**Key words:** Internet of Everything; edge computing; security requirements; privacy preserving

DOI: 10.12142/ZTETJ.201903006  
网络出版地址: <http://kns.cnki.net/kcms/detail/34.1228.TN.20190612.0847.002.html>

网络出版日期: 2019-06-12

收稿日期: 2018-12-28

## 1 万物互联背景下从云到边缘的演变

随着单片机嵌入式系统和无线通信技术的发展,物联网技术近年来得到了长足的发展和普及,其实际应用包括智能家居、智慧交通、无人驾驶等。近年来,除了“物”与“物”的互联,还增加了“物”与“人”的互联,其显著特点是“物”端具有更强的计算能力和语境感知

能力,将人和信息融入到互联网中,该趋势使得人类社会正在迈入万物互联(IoE)的时代<sup>[1]</sup>。

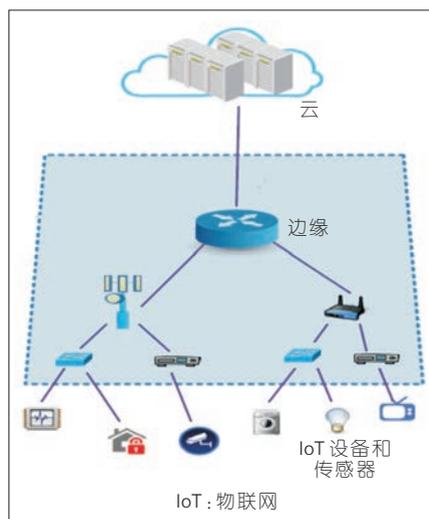
万物互联的核心在于收集来自于终端设备的海量数据,利用以大数据、机器学习、深度学习为代表的智能技术,去满足不同行业的业务需求,如制造、交通、医疗、农业等各行各业。在此背景下,所需要连接的终端设备数量达到数十亿甚至数万亿,其产生的数据呈爆炸式增

长。到2020年,连接到网络的无线设备数量将达到500亿台,生成的数据量达到507.9 ZB。

目前,海量数据的存储和处理主要依赖于集中式的云计算模型,其特征主要表现为数据和存储均位于部署在偏远地区的云计算中心。尽管云数据中心以堆叠硬件的方式具有较强的计算和存储能力,但是万物互联背景下,网络边缘的终端设备产生的数据已经达到海量级

别,这给云计算模型带来以下挑战:(1)线性增长的集中式云计算能力无法匹配终端所产生数据的指数增长需求<sup>[2]</sup>;(2)海量数据传输到云计算中心急剧增加了传输带宽的负载量,造成较大的网络时延,这给对时延敏感的应用场景(如无人驾驶、工业制造等)带来了严峻的挑战;(3)终端设备电能有限,数据传输会造成电能消耗较大。为此,集中式的云计算模型已经无法满足万物互联下的海量数据的高效传输以及处理需求。

在此背景下,边缘计算作为一种新的计算模式,架起物联网设备和数据中心之间的桥梁,使数据在源头附近就能得到及时有效地处理。如图1所示的基于物-边缘-云的三层服务交付架构,将从数据源到云计算中心数据路径之间的任意计算、存储、网络资源,形成高度虚拟化平台的“边缘层”为用户提供服务,其中的每层都具有灵活性和可扩展性,可以按需增减相应数量的实体。边缘计算出现之前,微云计



▲图1 物-边缘-云三层服务交付架构

算、雾计算和移动边缘计算等几种方法都是利用相似的思想为云计算提供了补充解决方案。根据2018年11月发布的《边缘计算参考架构3.0》所述:边缘计算模型具有分布式、“数据第一入口”、计算和存储资源相对有限等特性。

然而,网络边缘侧更贴近万物互联的终端设备。由于终端设备的开放性和异构性,以及相对有限的计算和存储资源(与云计算中心相比),使得访问控制和防护的广度和难度大幅提升<sup>[3]</sup>。此外,边缘计算还面临信息系统中普遍存在的网络攻击威胁。为此,跨越云计算和边缘计算之间的纵深,实施端到端的防护,全方位保障边缘计算的安全,增强其抵抗各种安全威胁的能力,是边缘计算促进万物互联进一步发展的前提和必要条件。

## 2 边缘计算安全需求

安全是指达到抵抗某种安全威胁或安全攻击的能力,横跨云计算和边缘计算,需要实施端到端的防护。万物互联系统在紧密耦合网络系统与物理世界中的关键性作用决定了安全属性和隐私保护的相关需求要比在以往任何信息系统中更加重要。

### 2.1 边缘计算安全的必要性

万物互联系统中终端设备具有超大规模、低成本设计、资源受限、设备异构等特性。同时,开发商重视功能优于安全的事实、用户更高的隐私要求、更难的信任管理使得保证万物互联系统的安全性显得更

具挑战性。

目前,边缘计算面临巨大的安全威胁。文献[4]中,作者分析了2个利用边缘计算应用场景的安全问题。一个是在智能制造工厂的场景下,攻击者可以篡改通信数据包,注入伪造的压力测量值欺骗决策器,延迟控制阀门的动作并造成设备损坏。如果没有适当的安全防范措施,不仅生产过程可能中断,工人的生命在很大程度上受到威胁。另一个则是移动边缘计算中无人机操作的安全问题,可以产生模拟的全球定位系统(GPS)信号误导无人机系统组件,使其驾驶到目标区域以达到捕获的目的。作者用无人机实验验证并达到了在不产生附加损失的情况下友好地捕获了非合作性无人机的目的。

在文献[5]中,华盛顿大学计算机科学与工程系的科研人员对配备智能电子控制系统的典型豪华轿车进行了实验安全分析,发现其安全保护系统具有很多设计时就存在的安全漏洞,同时并不是车上的所有组件都遵循其自己设计的安全协议,这使得实验人员能够轻易地侵入车辆引擎控制模块、电子刹车控制模块等性命攸关的重要车辆控制部件,从而远程控制行驶车辆的油门、刹车等。

2017年6月1日正式生效的《中华人民共和国网络安全法》特别强调了关键信息基础设施的运行安全,而能源、交通、制造等关键基础设施的工业控制环境无疑将是安全建设的重中之重。2016年中国信息通信研究院云计算白皮书指出:

公有云服务提供商向用户提供大量一致化的基础软件(如操作系统、数据库等资源),这些基础软件的漏洞将造成大范围的安全问题与服务隐患。安全已经成为阻碍万物互联和云计算发展的最大因素。

边缘计算是万物互联的延伸和云计算的扩展,三者的有机结合将为万物互联时代的信息处理提供较为完美的软硬件支撑平台,为能源、交通、制造、医疗等行业带来飞跃式发展。而通过边缘设备将类似云计算的功能带到了网络的边缘,可能引入新的安全挑战,一些传统的安全解决方法,例如基于非对称密钥协议和基于网际协议地址(IP)的解决方案,无法有效地应用于边缘计算系统,进而带来了一系列全新的安全需求。

同时,边缘计算可以提供理想的平台来解决物联网中的许多安全和隐私问题。在网络边缘处计算、连接、存储能力的协同使用可以达到万物互联应用的部分安全目的。例如,边缘设备可以作为加密计算的代理或者在公钥基础设施(PKI)技术中协助认证中心(CA)管理证书的发放与撤销,而其下面的物联网设备和传感器就缺乏实现这些操作的必要资源。

综上所述,网络边缘高度动态异构的复杂环境也会使网络难于保护,从而带来新的安全挑战。边缘计算同时又为资源、能量受限的终端设备提供一套全新的安全解决方案。因此,研究边缘计算场景的安全和隐私保护的相关问题是万物互联系统得到进一步发展的首要前提

条件。

## 2.2 边缘计算安全的需求分析

边缘计算安全需求分析如图2所示,按照边缘计算参考架构,主要分为物理安全、网络安全、数据安全和应用安全4个方面的需求。

### 2.2.1 物理安全需求

物理安全是保护智能终端设备、设施以及其他媒体避免自然界中不可抗力(如地震、火灾、龙卷风、泥石流)及人为操作失误或错误所造成的设备损毁、链路故障等使边缘计算服务部分或完全中断的情况。物理安全是整个服务系统的前提,物理安全措施是万物互联系统中必要且基础的工作。

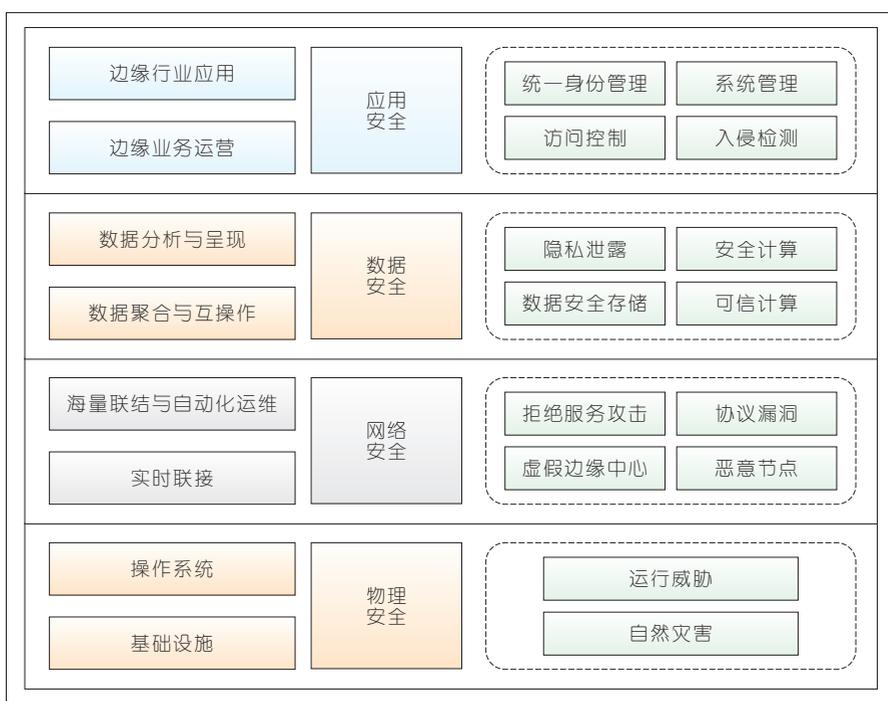
对于边缘计算设备来说,其在对外开放的、不可控的甚至人迹罕至的地方运行,所处的环境复杂多

样,因此更容易受到自然灾害的威胁。且在运行过程中,由间接或者自身原因导致的安全问题(如能源供应;冷却除尘、设备损耗等),运行威胁虽然没有自然灾害造成的破坏彻底,但是如果缺乏良好的应对手段,仍然会导致灾难性的后果,使得边缘计算的性能下降,服务中断和数据丢失。

### 2.2.2 网络安全需求

网络安全是指通过采用各种技术和管理措施,使网络系统正常运行,从而确保网络数据的可用性、完整性和保密性,以使系统连续可靠正常地运行,网络服务不中断。

大数据处理背景下,海量终端设备通过网络层实现与边缘设备的数据交互传输,边缘设备可以通过接入网络层实现更加广泛的互联功能。而大量设备的接入,给网络管



▲图2 边缘计算安全需求分析

理带来沉重负担的同时,也增加了边缘设备被攻击的可能性。文献[5-6]中所采用的攻击方式大都是在无线传输途中采用窃听、截获数据包等方法进行流量分析,然后篡改或伪造数据包来达到控制的目的。

相较于云计算数据中心,边缘节点的能力有限,更容易被黑客攻击。虽然单个被破坏的边缘节点损害并不大,并且网络有迅速找到附近可替代节点的调度能力;但如果黑客将攻陷的边缘节点作为“肉鸡”去攻击其他服务器,进而会对整个网络造成影响。现有大多安全保护技术计算保护流程复杂,不太适合边缘计算的场景。所以,设计适合于万物互联背景下边缘计算场景中轻量级的安全保护技术是网络安全的重大需求。

### 2.2.3 数据安全需求

数据信息作为一种资源,具有普遍性、共享性、增值性、可处理性和多效用性,而数据安全的基本目标就是要确保数据的3个安全属性:机密性、完整性和可用性。

要对数据的全生命周期进行管理的同时实现这3个安全属性才能保证数据安全。整个生命周期包括6个阶段<sup>[7]</sup>:创建,数据的产生和采集过程;存储,数据保存到存储介质的过程;使用,数据被浏览、处理、搜索或进行其他操作的过程;共享,数据在拥有者、合作者、使用者之间交互的过程;存档,极少使用的数据转入长期存储的过程;销毁,不再使用的数据被彻底删除和擦除的过程。

在边缘计算中,用户将数据外包给边缘节点,同时也将数据的控制权移交给边缘节点,这便引入了与云计算相同的安全威胁。首先,很难确保数据的机密性和完整性,因为外包数据可能会丢失或被错误地修改。其次,未经授权的各方可能会滥用上传的数据图谋其他利益。虽然相对于云来说边缘计算已经规避了多跳路由的长距离传输,很大程度地降低了外包风险;但是边缘计算设备部署的应用属于不同的应用服务商,接入网络属于不同的运营商,导致边缘计算中多安全域共存、多种格式数据并存。因此属于边缘计算的数据安全问题也日益突出,如在一个边缘节点为多个用户服务时,如何确保用户数据的安全隔离?在如此复杂多变的环境中,一个边缘节点瘫痪后,如何实现安全快速地迁移数据?当多个边缘节点协同服务时,如何能够在不泄露各自数据的情况下设计多方的协作服务?

另一个万物互联背景下边缘计算的数据安全需求就是用户隐私保护。比起云中心隐私数据泄露的风险,边缘计算设备位于靠近数据源的网络边缘侧,相对于位于核心网络中的云计算数据中心,可以收集更多用户高价值的敏感信息,包括位置信息、生活习惯、社交关系甚至健康状况等,边缘计算是否会成为商业公司收集用户隐私数据的平台?物联网设备的计算资源难以执行复杂的隐私保护算法,边缘式大数据分析中如何在数据共享时保证用户的隐私?这些问题都将成为边

缘计算发展的重要阻碍。

### 2.2.4 应用安全需求

应用安全,顾名思义就是保障应用程序使用过程和结果的安全。边缘式大数据处理时代,通过将越来越多的应用服务从云计算中心迁移到网络边缘节点,能保证应用得到较短的响应时间和较高的可靠性,同时大大节省网络传输带宽和智能终端电能的消耗。但边缘计算不仅存在信息系统普遍存在的共性应用安全问题,如拒绝服务攻击、越权访问、软件漏洞、权限滥用、身份假冒等,还由于其自身特性存在其他的应用安全的需求。在边缘这种多安全域和接入网络共存的场景下,为保证应用安全,该如何对用户身份进行管理和实现资源的授权访问则变得非常重要。身份认证、访问控制和入侵检测相关技术便是在边缘计算环境下保证应用安全的关键需求。

## 3 边缘计算安全挑战

通过对边缘计算安全需求的讨论分析可以看出:边缘计算的特性使其在构建安全保护方案时给系统开发人员带来了重大挑战。

### 3.1 身份认证

身份认证,也称“身份验证”或“身份鉴别”,是验证或确定用户提供的访问凭证是否有效的过程。用户可以是个人、应用或服务,所有的用户都应在被认证后才能访问资源,从而确定该用户是否具有对某种资源的访问和使用权限,使系统

的访问策略能够可靠、有效地执行,防止攻击者假冒合法用户获得资源的访问权限,保证系统和数据的安全,以及授权访问者的合法利益。

在边缘计算中,不同可信域中的边缘服务器、云服务提供商和用户分别提供和访问实时服务,其分散化、实时服务的低延迟需求和用户的移动性给身份认证的实现带来了巨大的障碍,很难保证所有涉及的实体都是可信的。在访问这些服务之前,应该对每个用户进行身份验证,以确保其真实性和可信性。身份认证应具备的功能包括:一方面应能够在分布式异构网络环境下,使用相关的协议、规范以及技术将分散的身份信息进行集中管理,实现单点登录,也可以方便地扩充跨身份标识域的访问等功能;另一方面应提供友好的体验环境,保护用户隐私,有效地对用户的行为进行审计。

身份认证是终端设备安全的基本要求,许多万物互联设备没有足够的内存和中央处理器(CPU)功率来执行认证协议所需的加密操作。这些资源有限的设备可以将复杂的计算和存储外包给可以执行认证协议的边缘设备,与此同时也会带来一定的问题:终端用户和边缘计算服务器之间必须相互认证,这种多安全域共存的情况下安全凭证从何而来?如何在大量分布式边缘服务器和云计算中心之间实现统一的身份认证和密钥管理机制?万物互联中存在大量的资源受限设备,无法利用传统的PKI体制对边缘计算设备或服务进行认证。边缘计算环境

下终端具有很强的移动性,如何实现用户在不同边缘设备切换时的高效认证具有很大挑战<sup>[8]</sup>。显然,轻量级的身份认证技术是保证边缘计算安全的前提和挑战。

### 3.2 访问控制

访问控制是基于预定模式和策略对资源的访问过程进行实时控制的技术,按用户身份及其所归属的某项定义组来限制用户对某些信息项的访问,或限制对某些控制功能的使用。访问控制的任务是在满足用户最大限度享受资源共享需求的基础上,实现对用户访问权限的管理,防止信息被非授权篡改和滥用,是保证系统安全、保护用户隐私的可靠工具。在万物互的联背景下,需要访问控制以确保只有受信任方才能执行给定的操作,不同用户或终端设备具有访问每个服务的独特权限。

访问控制除了负责对资源访问控制外,还要对访问策略的执行过程进行追踪审计。在边缘计算中,访问控制变得更加艰难,主要原因在于:首先要求边缘计算服务提供商能够在多用户接入环境下提供访问控制功能;其次,访问控制应支持用户基本信息和策略信息的远程提供,还应支持访问控制信息的定期更新;最后,对于高分布式且动态异构数据的访问控制本身就是一个重要的挑战。

### 3.3 入侵检测

入侵检测通过包括监测、分析、响应和协同等一系列功能,能够发

现系统内未授权的网络行为或异常现象,收集违反安全策略的行为并进行统计汇总,从而支持安全审计、进攻识别、分析和统一安全管理决策。从企业角度看,任何试图破坏信息及信息系统完整性、机密性的网络活动都被视为入侵行为。入侵检测技术广泛应用于云系统中,以减轻内部攻击、泛洪攻击、端口扫描、虚拟机攻击和hypervisor攻击等入侵行为。

在边缘计算中,外部和内部攻击者可以随时攻击任何实体。若没有实施适当的入侵检测机制来发现终端设备和边缘节点的恶意行为或协议违规,则会逐步破坏服务设施,进而影响整个网络。

但是,在万物互联环境下,由于设备结构、协议、服务提供商的不同,难以检测内部和外部攻击<sup>[9]</sup>。此外,如何通过资源能力受限的边缘设备间的系统来进行全局的入侵检测,使其能够在大规模、广泛地理分布和高度移动的环境中得到应用,具有十分重要的意义。

### 3.4 隐私保护

万物互联系统的目标是通过收集海量数据为用户提供多种个性化服务。由于终端设备资源受限,缺乏对数据加密或解密的能力,这使得它容易受到攻击者的攻击。

边缘计算将计算迁移到临近用户的一端,直接对数据进行本地处理、决策,在一定程度上避免了数据在网络中长距离的传播,降低了隐私泄露的风险。然而,由于边缘设备获取的用户第一手数据,能够获

得大量的敏感隐私数据。如何能够保证用户在使用服务的同时又不泄露其敏感信息对边缘计算中的隐私保护算法提出了更高的要求。

### 3.5 密钥管理

密钥管理包括从密钥产生到密钥销毁的各个方面,主要表现于管理体制、管理协议和密钥的产生、分配、更换和注入等,包含密钥生成、密钥分发、验证密钥、更新密钥、密钥存储、备份密钥、密钥的有效期、销毁密钥这一系列的流程。密钥在已授权的加密模块中生成,高质量的密钥对于安全是至关重要的,整个密码系统的安全性并不取决于密码算法的机密性,而是取决于密钥的机密性。一旦密钥遭受泄露、窃取、破坏,机密信息对于攻击者来说已经失去保密性。由此可见,密钥管理对于设计和实施密码系统而言至关重要。

在万物互联环境中,由于云服务商、边缘服务商和用户对密钥管理系统与信息技术(IT)基础设施具有不同的所有权和控制权,这使得面向边缘计算环境的密钥管理比传统信息系统的密钥管理更为复杂。因为每个应用出于特定的安全目的管理其安全密钥,使得跨应用密钥管理变得尤为复杂,参与多个应用程序的用户设备需要管理多个安全密钥或口令,增加了密钥泄露的风险并危及服务的安全性。显然,在大规模、异构、动态的边缘网络中,保证用户和用户之间、用户和边缘设备之间、边缘设备和边缘设备之间、边缘设备和云服务器之间的信

息交互安全,给边缘计算模式下实现高效的密钥管理方案带来了严峻的挑战。

### 4 结束语

随着万物互联时代的到来,基于云计算模型的集中式大数据处理模式已经无法满足网络边缘设备所产生海量数据处理的实时性、安全性和低功耗等需求。为此,将原有云计算中心的部分或者全部计算任务迁移到数据源的附近执行,边缘计算在梯联网、工业机器人、无人驾驶、智慧交通等领域扮演着越来越重要的角色。作为一种新型的去中心化架构,它将云计算的存储、计算和网络资源扩展到网络边缘,以支持大规模的协同万物互联应用。然而,由于边缘设备更加靠近网络边缘侧,网络环境更加复杂,并且边缘设备对于终端具有较高的控制权限,导致其在提高万物互联网络中数据传输和处理效率的同时,不可避免地带来一些新的安全威胁,如物理安全、网络安全、数据安全、应用安全等。同时,边缘计算模式也给身份认证、访问控制、入侵检测、隐私保护、密钥管理等方面带来了严峻的挑战。为此,我们需要清晰地认识边缘计算安全框架和业务流程,设计安全的边缘计算架构,这些对于促进边缘计算的进一步普及和发展具有十分重要的意义。

#### 参考文献

[1] 施巍松, 孙辉, 曹杰, 等. 边缘计算: 万物互联时代新型计算模型[J]. 计算机研究与发展, 2017,

54(5): 907-924. DOI:10.7544/issn1000-1239.2017.20160941  
 [2] 施巍松, 刘芳, 孙辉, 等. 边缘计算[M]. 北京: 科学出版社, 2018  
 [3] 边缘计算产业联盟, 工业互联网产业联盟. 边缘计算与云计算系统白皮书[R]. 2018  
 [4] HE D J, CHAN S, GUIZANI M. Security in the Internet of Things Supported by Mobile Edge Computing [J]. IEEE Communications Magazine, 2018, 56(8): 56-61. DOI:10.1109/mcom.2018.1701132  
 [5] KOSCHER K, CZESKIS A, ROESNER F, et al. Experimental Security Analysis of a Modern Automobile[C]//2010 IEEE Symposium on Security and Privacy. USA: IEEE, 2010: 447-462. DOI:10.1109/SP.2010.34  
 [6] 云计算白皮书[EB/OL]. [2018-12-20]. http://www.199it.com/archives/764250.html  
 [7] 卿昱. 云计算安全技术[M]. 北京: 国防工业出版社, 2016:53-54  
 [8] HE D B, ZEDADALLY S, WU L B, et al. Analysis of Handover Authentication Protocols for Mobile Wireless Networks Using Identity-Based Public Key Cryptography [J]. Computer Networks, 2017, 128: 154-163. DOI:10.1016/j.comnet.2016.12.013  
 [9] PRABAVATHY S, SUNDARAKANTHAM K, SHALINIE S M. Design of Cognitive Fog Computing for Intrusion Detection in Internet of Things[J]. Journal of Communications and Networks, 2018, 20(3): 291-298. DOI:10.1109/jcn.2018.000041

#### 作者简介



马立川, 西安电子科技大学网络与信息安全学院师资博士后; 主要研究方向为隐私保护、边缘计算安全等。



裴庆祺, 西安电子科技大学教授、博士生导师, 综合业务网理论与关键技术国家重点实验室(ISN)成员、区块链应用与评测研究中心主任, 西安市移动边缘计算及安全重点实验室主任, 陕西省区块链与安全计算重点实验室执行主任; 主要研究方向为认知网络与数据安全、区块链、边缘计算及安全等领域。



肖慧子, 西安电子科技大学通信工程学院在读博士研究生; 主要研究方向为边缘计算和物联网中的安全与隐私问题。

# 自然环境下基于异构多源的边缘计算公共安全系统

## Edge Computing-Based Systems for Multi-Source Heterogeneous in Public Safety

孙辉/SUN Hui, 余莹/YU Ying

(安徽大学, 安徽 合肥 230031)  
(Anhui University, Hefei 230031, China)



**摘要:** 现有基于集中式大数据的公共安全系统较难满足实时性、精准性、隐私性及能耗等问题。基于边缘计算的多源大数据的时代已经到来,多源数据主要包括:社会中政务和媒体数据、各种传感器收集的数据、智能终端产生的自媒体数据等。边缘计算与传统集中式云计算二者相结合,可较好地处理这些边缘异构多源数据。提出了基于边缘计算的智能交通系统、基于边缘计算的公共安全监测控制系统以及基于边缘计算的灾难预警与救援系统。

**关键词:** 边缘计算;多源异构;智能终端;协同

**Abstract:** Currently, cloud computing-based public safety systems are challenging to meet the requirements of real-time, accuracy, privacy, and energy consumption. The multi-source and heterogeneous-based big data applications that use the edge computing paradigm has become the popular applications. Multi-source-based big data mainly includes: (1) big data for governments and social media data; (2) data collected by various sensors; (3) self-media data from intelligent end devices. Edge computing, which collaborates with traditional cloud computing, can better handle these edge heterogeneous multi-source data. In this paper, three types of edge computing-based systems for heterogeneous multi-source are proposed: an edge computing-enabled intelligent transportation system, an edge computing-based monitoring system in public safety, and a disaster alert and notification system using the edge computing paradigm.

**Key words:** edge computing; multi-source heterogeneous; intelligent device; collaboration

DOI: 10.12142/ZTETJ.201903007  
网络出版地址: <http://kns.cnki.net/kcms/detail/34.1228.TN.20190614.0934.001.html>

网络出版日期: 2019-06-14  
收稿日期: 2018-12-10

## 1 背景

### 1.1 边缘计算

随着物联网技术的发展,网络边缘设备数量的快速增加,万物互联时代已经到来。现有公共安全系统主要建立在云计算模式

下,随着其规模的扩大,监测控制视频数据的处理难以满足实时性需求;而基于边缘计算的公共安全系统正逐渐成为主流,监测控制数据由原有的单一视频转变为以声音、视频、文本等多源数据。

边缘计算指在网络边缘执行计算的一种新型计算模型,边缘计算中的边缘指从数据源到云计算中心路径间的任意计算和网络资源<sup>[1]</sup>。

边缘大数据处理时代,物联网设备的数据急剧增加,数据类型呈现出复杂化、多样化的特征。我们正从以云计算为核心的集中式大数据处理时代步入异构多源的边缘式大数据处理时代。

相比于云计算模型的集中式大数据处理时代,万物互联下的边缘式大数据具有异构多源的特征。云计算时代,数据量通常维持在拍字

基金项目:安徽省社科规划青年项目(AHSKQ2015D108);安徽省重点研究与开发计划项目(1704d0802193)

节(PB)级别,而边缘大数据所处理的数据量已超过泽字节(ZB)。根据思科全球云指数的预测:至2020年,全球数据中心的流量将达到15.3 ZB<sup>[2]</sup>。由于基于边缘数据的应用程序对实时性需求较高,原有云中心的计算任务被(部分或全部)迁移到边缘计算单元上,以保证数据处理的实时性<sup>[1]</sup>,同时降低数据从边缘设备传输到云中心的带宽负载。因此,边缘计算架构目前已经逐渐成为异构多源大数据时代的重要架构之一。

## 1.2 公共安全与边缘大数据

根据数据来源的不同,公共安全领域<sup>[3]</sup>可利用的边缘大数据包括:政务和媒体数据、传感器收集的物理空间的数据和网络空间的数据等<sup>[4]</sup>。政务和媒体数据主要包括政府数据、传统媒体数据等,政府通过其行政权、司法权及强制力量可收集大量公共安全相关数据,传统媒体(如电视台)依靠其专业性,在获取数据上也有其独特的优势。物理空间数据主要包括公共场所的摄像头监测控制数据、车载摄像头拍摄到的移动视频数据,环境传感器收集到的温度、压力、湿度数据,烟雾传感器收集的室内烟雾数据。这些来自于物理空间的数据均可应用到公共安全领域,有效融合这些数据,是公共安全系统的难题之一。网络空间数据主要包括自媒体数据,如用户通过 Facebook<sup>[5]</sup>、Twitter<sup>[6]</sup>、微博、微信等社交平台分享照片及视频;日志数据,如用户通过浏览器观看在线视频数据、搜索引擎资源、网

上银行交费等行为数据。网络数据是异构多源数据重要组成部分,据统计,YouTube<sup>[7]</sup>用户每分钟上传视频内容时长达72 h;Twitter用户每分钟达近30万次的访问量;Instagram<sup>[8]</sup>用户每分钟上传近22万张新照片<sup>[9]</sup>。网络空间数据是异构多源公共安全系统的主要来源。如当灾难发生后,通过社交网络上的数据分析灾民的行为模式,对受灾城市可能发生的大规模灾民避难、迁移行为进行建模、预测和模拟,对灾民的撤离路线进行有效推荐<sup>[10]</sup>。

## 1.3 现有公共安全系统特征及问题

目前,物联网技术在视频监控控制、城市交通管理等公共安全领域得到广泛应用;但传统的公共安全系统仅从传感器收集数据,数据来源较单一。物联网设备的计算能力有限,联网资源共享和智能化应用水平较低,仍处在物联网应用的初级阶段<sup>[11]</sup>。如2016年二十国集团峰会(G20)期间<sup>[12-13]</sup>,杭州采取了最高级别安保措施以保障峰会的顺利举办;但将安全系统扩大到整个城市时,其自动化能力不足,城市的安全监测控制主要依靠人力,杭州为此出动的警力达三万余人。此外,边缘端设备所产生的数据量越来越大,网络带宽<sup>[14]</sup>以及云中心的计算资源逐渐达到瓶颈。这限制了传统公共安全系统规模的扩展,并且实时性较低,严重影响了公共安全系统对突发事件的及时处理。

本文中,我们提出利用异构多源数据来构建基于边缘计算的公共

安全系统。边缘计算单元可及时处理由大量异构传感器采集的海量数据<sup>[15]</sup>,在靠近数据源的一端,对数据进行预处理,过滤无用数据,降低传输带宽,将有用信息传输到云中心。系统自动收集网络空间数据、政府数据,对多源数据进行融合,从而为大规模城市公共安全系统的构建提供了可能。

## 2 异构多源公共安全系统的挑战

我们提出自然环境下基于异构多源的边缘计算公共安全系统,该类系统的关键技术挑战主要包括以下几个方面。

### 2.1 边缘计算编程模型

异构多源公共安全系统的边缘端大多是异构平台,缺少统一化的编程模型,这给程序员开发边缘端软件带来了阻力。在云计算模型中,程序员编写应用程序并将其部署到云端,云服务器提供商维护云计算服务器,基础设施对程序员是透明的。但是,边缘计算模型从云端迁移部分或全部计算任务到边缘端,因为每个节点的运行时环境可能不尽相同,传统编程方式 MapReduce<sup>[16]</sup>、Spark<sup>[17]</sup>均不适合,这对程序员困难较大。

### 2.2 数据抽象

作为数据生产者的传感器向边缘计算单元发送数据,计算单元对数据进行预处理(如去除无用信息、事件检测、隐私保护等),之后数据被传送到云端。由于边缘传感器收

集的数据具有异构多源的特性,数据格式具有多样性,这会直接影响数据的使用。由于数据抽象会导致大量信息的损失,如何保留有用信息、保证数据抽象的轻量性及有用信息的高度浓缩,是数据抽象的一个难点。

### 2.3 任务调度策略

在异构多源公共安全系统中,数据、计算、存储、网络等资源具有异构性,须要针对不同应用实例设计基于边缘计算的异构资源任务调度策略<sup>[18]</sup>。大多应用程序具有多样性,调度策略应支持多种类型的应用程序。但是,在边缘计算架构下,计算资源受限,利用基于边缘服务器和边缘计算节点的分布式<sup>[19]</sup>系统以及有限的计算资源<sup>[20]</sup>,从多角度卸载<sup>[21]</sup>和分发子任务到资源合适的边缘节点,降低资源占用,提高执行效率,是边缘计算公共安全系统所面临的重要挑战之一。

### 2.4 信息安全

在内容感知、实时计算、并行处理等方面,基于异构多源的边缘计算公共安全系统具有开放性,这使得数据安全与隐私问题变得更加突出<sup>[22]</sup>。R.ROMAN等<sup>[23]</sup>对几种移动边缘范式进行了安全性分析,阐述了一种通用协作的安全防护体系,这些工作为边缘计算的安全研究提供了理论借鉴。数据隐私保护及安全是异构多源公共安全系统提供的一种重要服务,如基于边缘计算的城市安全视频监控控制系统,大量的隐私信息被摄像机捕获,在隐私

保护下提供服务,这对异构多源的公共安全系统又一挑战。

### 2.5 大规模系统运维

随着边缘数据呈现异构多源化,系统运维<sup>[24]</sup>将会成为大型公共安全系统的一个巨大挑战。异构多源公共安全系统与普通系统不同,其在数据规模上是城市级别的,数据来源于城市的公共摄像头、移动车载摄像头、公共网络、政府数据等,一个800像素的摄像机每小时产生的数据量是3.6GB,即使经过边缘单元的数据抽象压缩,数据量仍然较大。要维护如此庞大的异构多源数据系统,对运维工作者来说也是一个巨大的挑战。

## 3 面向异构多源数据的3种边缘计算公共安全系统

基于公共安全系统的需求,以及边缘计算的适用场景,我们在以下3个方面提出关于公共安全系统的展望。

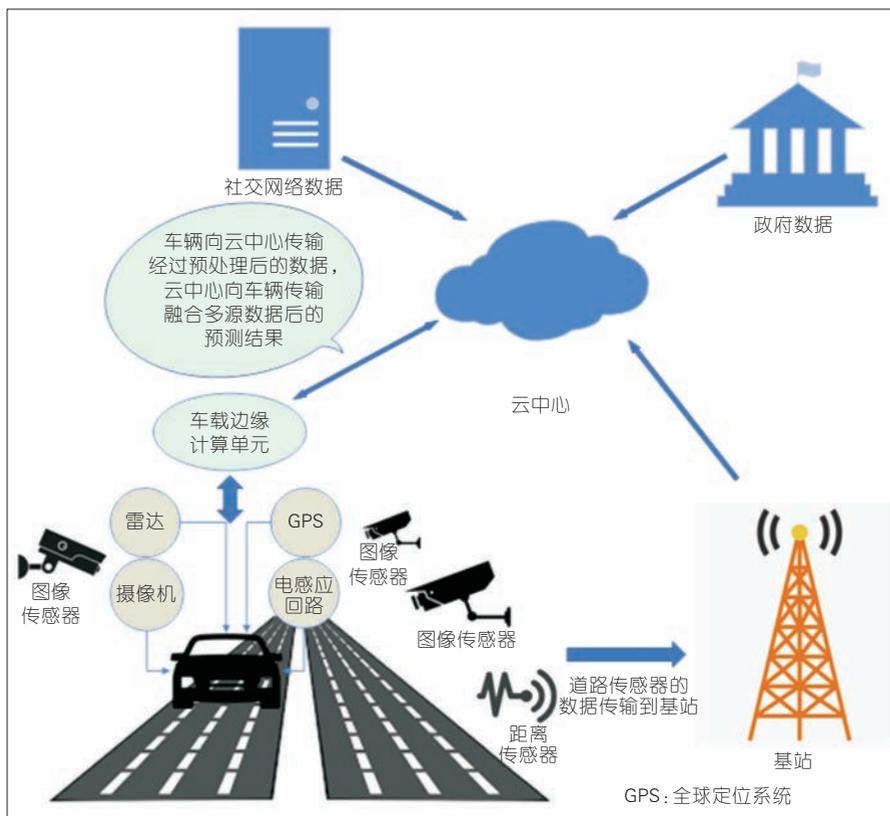
### 3.1 基于边缘计算的智能交通系统

智能交通是一种将先进通信技术与交通技术相结合的物联网重要应用。智能交通用于解决城市居民面临的出行问题,如恶劣的交通现状、拥堵的路面条件、贫乏的停车场等<sup>[25]</sup>。智能交通系统可以利用车辆及车辆周边的传感器网络进行交通信息的交换<sup>[26]</sup>,并主动做出决策,同时会实时收集来自社交网络和政府的数据,预测未来前方车流量,并进行规避。随着交通数据量的增

加,用户对交通信息的实时性需求也在提高。若将所有数据传输到云中心,将造成带宽浪费和延时,也无法优化基于位置识别的服务。基于异构多源的边缘计算智能交通为上述诸多问题提供了一种较好的解决方案,是边缘计算公共安全系统重要组成部分之一。

安全性是智能交通行业最重要的问题。基于异构多源边缘计算智能交通系统能够提高智能交通的安全性。若将车载传感器实时数据全部传输到云端,并将云端反馈结果用于对车辆的控制,将会造成极大的延迟。在无人驾驶或自动驾驶<sup>[27]</sup>领域,这些平时看上去不起眼的延迟在某些突发事件发生时,往往会造成灾难。为此,智能交通系统加入了边缘计算模块,将车载传感器收集到的数据进行本地化处理,并抽象出核心数据,使其符合边缘计算编程模型,并将其传输到云端处理。这不仅能节省带宽和能耗,同时能降低网络延时,增强系统的实时处理能力。同时,智能交通系统将社交网络和政府数据也传输到云端处理,用于预测未来一段时间内可能发生的突发车流变化,如节假日、大型会议、救护车、警车等。有效调度异构多源数据并辅以边缘计算,能够极大提高智能交通系统的公共安全性,同时能对智能交通系统进行合理、及时的运维,也是对公众安全一种保障。

如图1所示,利用异构多源数据(电感应回路探测器、雷达、摄像机、众包、社交媒体等),通过深度学习<sup>[28]</sup>的方式预测出了未来交通流<sup>[29]</sup>



▲ 图1 基于边缘计算的智能交通系统架构图

是智能交通数据处理的一种趋势。本系统中的异构多源数据不仅包括传感器的信息,还包括其他来源数据,对这些数据资源进行融合分析,将人们的出行意愿和突发事件进行数据挖掘<sup>[30]</sup>,可以实现对未来的交通流进行更深层次、更精准的预测。如在节假日期间,因为历史数据未能考虑交通流突变的情况,数据分析模型将很可能无法做到良好的预测;但结合网络空间的数据,收集人们在社交网络空间上分享的出行意愿,并实时处理,就可以避免车流量高峰,预测出低流出行路线,这对人们在节假日的出行安全提供了良好保障。再如,将政府发布的信息(如大型会议、灾难性事件),进行多源融合,通过深度神经网络进行

实时分析,将实现对交通流量的预测,保证智能安全出行。

### 3.2 基于边缘计算公共安全监测控制系统

城市安全监测控制系统广泛应用于新型犯罪及社会管理等公共安全问题。传统视频监测控制系统前端摄像头内置计算能力较低,而现有智能视频监测控制系统的智能处理能力不足。传统的大型监测控制系统将摄像头采集的数据简单压缩或直接传输到云中心,而随着监测控制系统的规模不断扩大,公共安全监测控制领域要求系统能够提供实时性较高的数据。

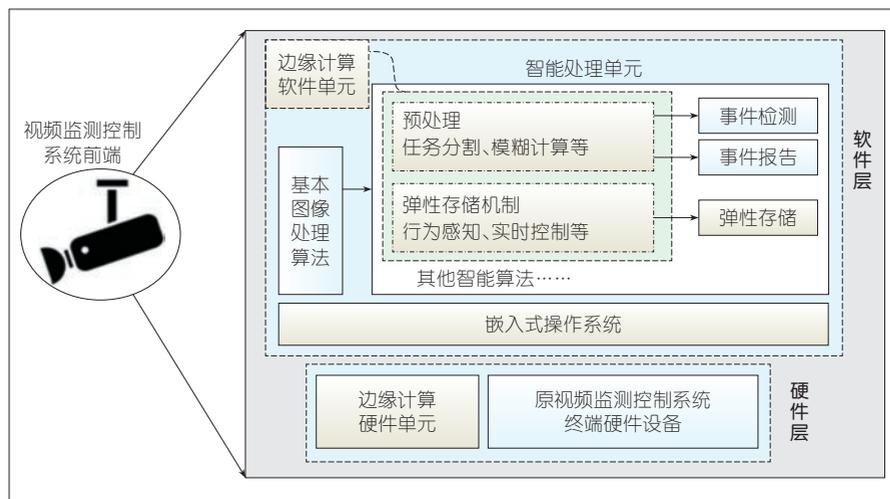
从异构多源融合的角度,我们提出边缘计算公共安全系统的新架

构,如加入智能手机摄像头收集的数据、网友发布的实时求救信息,以及政府发布的搜查公告、红色通缉令等。

在边缘端,边缘计算下的公共安全监测控制系统发展趋势是以云计算和物联网技术为基础,融合边缘计算架构和异构多源数据,将有边缘计算能力的硬件集成到传统摄像头上,辅以软件技术,构建基于边缘计算的新型监测控制应用的软硬件服务平台。图2为基于边缘计算的视频监测控制系统框图,其中具有边缘计算功能的模块作为协处理单元(简称边缘计算硬件单元)与原有视频监测控制系统的摄像头终端系统进行系统融合<sup>[25]</sup>。

针对大型监测控制系统故障检测、视频监测控制系统内容可用性、网络带宽有限、云中心负载重等问题,提高监测控制系统的智能处理能力,进而实现重大刑事案件和恐怖袭击活动预警系统和处置机制,提高公共安全监测控制的防范刑事犯罪和恐怖袭击的能力,降低视频数据网络传输和云中心负载。

在云中心,本系统收集网络空间和政务信息的数据并分析,运用深度学习与大数据处理技术,抽象出需要监测控制对象的特征,并与从边缘端传到云端的数据进行比对,达到自动化多目标安全监测控制的目的。该系统自动调度从政府与网络数据中提取的、可能会对公共安全造成危害的敏感人员信息,通过边缘计算视频监测控制系统,按照边缘计算编程模型所需的格式将数据下发给公共摄像头和车载摄



▲图2 基于边缘计算视频监控控制终端

摄像头,在城市范围内对敏感人员进行目标识别,并监测控制其行为。通过政府数据获得嫌疑人的社会关系网络,对嫌疑人及其相关人员通过人脸识别<sup>[21]</sup>进行监测控制,同时对嫌疑人的不动产和名下汽车车牌进行搜查。此外,通过社交网络上的信息对可疑区域进行重点监测控制等。一旦边缘计算单元检测出其进行或蓄意进行危害公共安全的行为,就将监测控制数据传送给云端,利用云端进行广播<sup>[22]</sup>报警。

相比于传统监测控制方案,边缘计算公共安全监测控制系统可提升自动化程度,降低对控制中心的人工依赖,提高信息获取的实时性。同时,边缘计算架构缓解了视频数据上传至云端所带来的带宽负载和传输延迟问题,这将极大加强公共安全系统的监测控制能力,为公共安全提供更深层的保障。

### 3.3 基于边缘计算的灾难预警与救援系统

在科技高速发展的同时,人类

依然无法在地震、水灾等自然灾害发生前做好精准性预测。在灾难现场,通信网络(如固定网络、无线通信网络等)往往不稳定且鲁棒性差,如何基于有限的网络环境,将信息高效地传输至云端是实现高效、实时救灾系统的关键之一。基于无人机、视频、众包等方式在灾难环境下收集信息已成为趋势。灾难现场传感器传回的视频、音频、社交网络等数据有助于对受灾后的灾区环境、人群、救灾通道等进行智能感知,在灾难响应和恢复救援工作中发挥着越来越重要的作用。

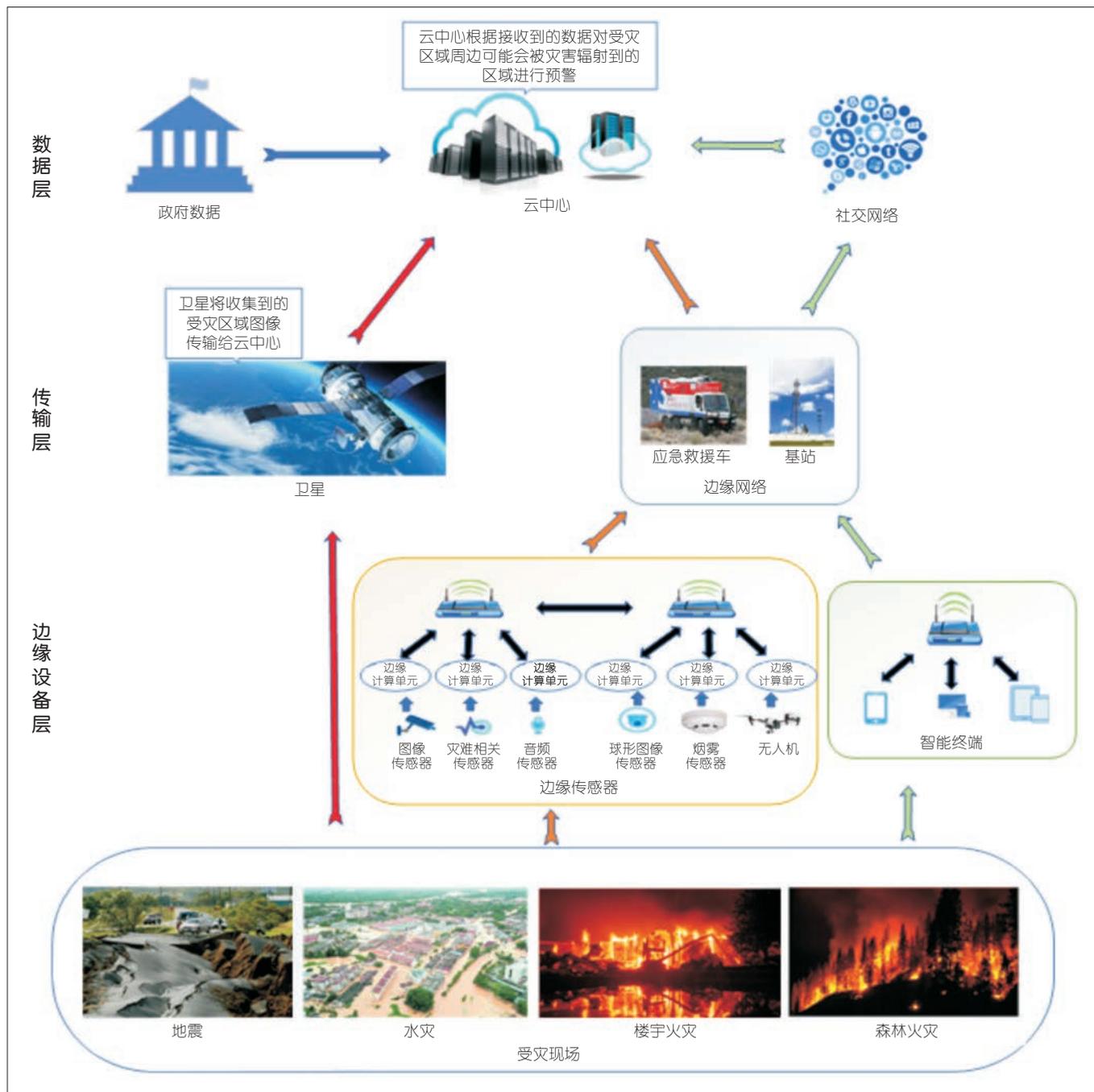
现有自然灾害实时预测方法大体上是通过检测自然灾害产生时的某项指标,如通过电磁波、天气状况、地震云、土壤或地下水的氦气或氢气含量、井水水位及动物行为等预测地震的发生,这种思路具有很强的争议性。本文中,我们提出的基于边缘计算的灾难预警系统,可以在某个区域发生灾难时,通过收集这个区域内边缘传感器的数据(如光学传感器、红外摄像机、麦克

风等)和网络上的数据,进行边缘大数据分析,实时准确地判断出灾难发生的具体类型及辐射范围,自动地将分析结果辐射到相邻区域。在一定程度上实现了对将要发生灾难的相邻区域的预警,为应急响应和搜救行动的完成获取相关信息。

本文中我们构想的预警系统,通过边缘传感器收集受灾发生现场的实时数据,如图3所示。本文所提出的灾难救援系统将从无人机、智能机器人、灾难现场固定传感器收集来的图像、音频等传感器数据(如烟雾传感器数据、温度传感器数据等),通过边缘计算模块进行内容感知和数据抽象,将处理之后的数据上传至云中心,云中心再将这些传感器的信息与卫星拍摄的卫星灾情图片、灾难现场上传至社交网络上的信息等多源数据进行实时处理,对受灾区域周边可能会被灾害辐射到的区域进行预警,避免灾害的扩大。

当灾难发生时,人们因为焦虑、纪念、好奇等原因会在网络空间发布信息,这种信息本质上是一种信息、观点、技术、情绪和行动的聚合。这就产生了社会边缘大数据,为灾难环境下,通过大数据分析人们的行为信息提供了可能。此外,政府通过社交媒体推送灾难的相关信息,网民可以发表自己的评论,这个过程又会促进社会边缘大数据的扩大。

灾难发生后,如何对受灾群众进行有效、及时的救援是公共安全系统要考虑的另一个问题。在危机发生后,网络传输环境复杂,网络传



▲ 图3 基于边缘计算的灾难预警与救援系统

输不稳定。若将灾难现场传感器所有信息(如视频、音频等)上传至云中心,网络性能将成为信息收集的瓶颈,并影响其他实时救灾任务的进行。基于边缘计算的灾难救援系统可以实现传感器异构多源数据的

本地化处理,主动式地适应灾难区域的网络传输条件,合理调度资源,将数据高效、安全、实时地上传至云中心,降低了极端网络环境下传输的数据量,减少云中心计算资源消耗,实现灾情数据的及时处理。

云中心将从社交网络上收集到的信息与灾难区域发送来的信息进行多源融合,比对、匹配出救灾线索,为救灾行动提供指导,进而制定更为安全可行的营救方案。救灾系统的稳定性决定着救灾的效率,合

理规划系统运维工作是公共安全系统的保障。

## 4 结束语

随着对公共安全系统功能以及规模要求的提高,边缘传感器数量增加,我们已经进入边缘大数据时代。传统的云计算架构无法在规模、功能和实时性上满足公共安全系统的需求,同时传统公共安全系统无法有效解决异构多源的数据融合、云中心负载较重、传输带宽较低以及数据隐私等问题。本文中,我们阐述了公共安全系统的需求,分析了在边缘大数据时代复杂网络结构下边缘大数据的来源,对基于异构多源的边缘计算公共安全系统进行展望,并提出3种系统架构:基于边缘计算的智能交通系统、基于边缘计算的公共安全检测控制系统、基于边缘计算的灾难预警与救援系统。物联网在智慧城市建设中受到学术界和产业界的高度重视,边缘计算的兴起为物联网增添了智能感知的能力。异构多源公共安全系统的发展将越来越受到重视,而边缘计算这一关键技术将在其中发挥极其重要的作用。

## 致谢

感谢清华大学郑纬民教授对本文相关研究的帮助和指导。

## 参考文献

- [1] 施巍松,孙辉,曹杰,等.边缘计算:万物互联时代新型计算模型[J].计算机研究与发展,2017,54(5):907.DOI: 10.7544/issn1000-1239.2017.20160941
- [2] Cisco Cloud Index Supplement. Cloud Readiness Regional Details White Paper[R]. 2017

- [3] 中国公共安全科技问题分析与发展战略规划研究. 中国工程科学, 2007, 9 (4): 35-40
- [4] 郭贺铨. 大数据思维[J]. 科学与社会, 2014, 4(1): 1. DOI:10.19524/j.cnki.10-1009/g3.2014.01.001
- [5] Company Info | Facebook Newsroom [EB/OL]. [2018-12-01]. <https://newsroom.fb.com/company-info/>
- [6] Twitter - Company [EB/OL]. [2018-10-01]. <https://about.twitter.com/zh-cn/company.html>
- [7] YouTube Architecture - High Scalability [EB/OL]. [2018-12-01]. <http://highscalability.com/blog/2008/3/12/youtube-architecture.html>
- [8] Instagram APKs - APKMirror [EB/OL]. [2018-12-01]. <https://www.apkmirror.com/apk/instagram/instagram-instagram/>
- [9] Data Never Sleeps 2.0 | Domo [EB/OL]. [2017-10-09][2018-12-01]. <https://www.domo.com/learn/data-never-sleeps-2/2017-10-09>
- [10] 宋轩. 大数据下的灾难行为分析和城市应急管理[J]. 中国计算机学会通讯, 2013, (8)
- [11] 刘攀, 赵晓艳, 李凯丽. 物联网测控技术在公共安全领域的发展与应用[J]. 中国新通信, 2017, 19(22): 27. DOI:10.3969/j.issn.1673-4866.2017.22.021
- [12] 林孝平. 上海世博会园区安全控制系统建设综述[J]. 电信快报, 2013, (9): 10
- [13] 浙江杭州获得2016年G20峰会举办权. 网易新闻[EB/OL]. [2015-02-28][2018-12-01]. <http://news.163.com/15/0228/22/AJSSJ1000145EH.html>
- [14] BUYYA R, YEO C S, VENUGOPAL S, et al. Cloud Computing and Emerging IT Platforms: Vision, Hype, and Reality for Delivering Computing as the 5th Utility [J]. Future Generation Computer Systems, 2009, 25(6): 599. DOI:10.1016/j.future.2008.12.001
- [15] ORSINI G, BADE D, LAMERSDORF W. Computing at the Mobile Edge: Designing Elastic Android Applications for Computation Offloading[C]//2015 8th IFIP Wireless and Mobile Networking Conference (WMNC). USA: IEEE, 2015. DOI:10.1109/wmnc.2015.10
- [16] DEAN J, GHEMAWAT S. MapReduce: Simplified Data Processing on Large Clusters [J]. Communications of the ACM, 2008, 51(1): 107-113
- [17] ZAHARIA M, CHOWDHURY M, FRANKLIN M J, et al. Spark: Cluster Computing with Working Sets[C]//Proceedings of the 2nd USENIX Conference on Hot Topics in Cloud Computing. USA: USENIX Association Berkeley, 2010
- [18] Fitzsimmons J A, Fitzsimmons M J. Service Management: Operations, Strategy, and Information Technology [M]. USA: McGraw-Hill Higher Education, 2004
- [19] COULOURIS G, DOLLIMORE J, KINDBERG T, et al. Distributed Systems: Concepts and Design(5th Edition) [M]. USA: Pearson, 2011
- [20] CHAITIN G J. On the Length of Programs for Computing Finite Binary Sequences: Statistical Considerations [J]. Journal of the ACM, 1969, 16(1): 145. DOI:10.1145/321495.321506
- [21] KUMAR K, LU Y H. Cloud Computing for Mobile Users: Can Offloading Computation Save Energy? [J]. Computer, 2010, 43(4): 51. DOI:10.1109/mc.2010.98
- [22] 张佳乐, 赵彦超, 陈兵, 等. 边缘计算数据安全与隐私保护研究综述[J]. 通信学报, 2018, 39(3): 1. DOI:10.11959/j.issn.1000-436x.2018037
- [23] ROMAN R, LOPEZ J, MAMBO M. Mobile Edge Computing, Fog et al.: A Survey and Analysis of Security Threats and Challenges [J]. Future Generation Computer Systems, 2018, 78: 680. DOI:10.1016/j.future.2016.11.009
- [24] O'BRIEN J A. Management Information Systems: Managing Information Technology in the Networked Enterprise [M]. USA: McGraw-Hill Professional, 1998
- [25] 施巍松, 刘芳, 孙辉. 边缘计算[M]. 北京: 科学出版社, 2018
- [26] DIMITRAKOPOULOS G, DEMESTICHAS P. Intelligent Transportation Systems [J]. IEEE Vehicular Technology Magazine, 2010, 5(1): 77. DOI:10.1109/mvt.2009.935537
- [27] COHEN S. Self-Driving Car [J]. Science and Children, 2017, 54(6): 96
- [28] LECUN Y, BENGIO Y, HINTON G. Deep Learning [J]. Nature, 2015, 521(7553): 436. DOI:10.1038/nature14539
- [29] LV Y, DUAN Y J, KANG W W, et al. Traffic Flow Prediction with Big Data: A Deep Learning Approach [J]. IEEE Transactions on Intelligent Transportation Systems, 2014: 1. DOI:10.1109/tits.2014.2345663
- [30] HAND D J. Principles of Data Mining [J]. Drug Safety, 2007, 30(7): 621. DOI:10.2165/00002018-200730070-00010
- [31] TURK M A, PENTLAND A P. Face Recognition Using Eigenfaces[C]//Proceedings. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. USA: IEEE, 1991. DOI:10.1109/cvpr.1991.139758
- [32] FALL K R, STEVENS W R. TCP/IP Illustrated, Volume1: The Protocols(2nd Edition) (Addison-Wesley Professional Computing Series)[M]. USA: Addison-Wesley Professional, 2011

## 作者简介



孙辉, 安徽大学讲师; 从事计算机系统、数据存储系统、边缘计算、基于数据的计算、GPGPU高性能数据处理等方面的工作; 已发表论文10余篇。



余莹, 安徽大学计算机科学与技术学院在读硕士研究生; 主要研究方向为计算机系统、边缘计算及边缘适用系统。

# 工业互联网智能制造 边缘计算：现状与挑战

## Industrial Internet Intelligent Manufacturing Edge Computing: State-of-the-Art and Challenges



宋纯贺/SONG Chunhe<sup>1,2</sup>, 曾鹏/ZENG Peng<sup>1,2</sup>, 于海斌/YU Haibin<sup>1,2</sup>

(1. 中国科学院沈阳自动化研究所, 辽宁 沈阳 110016;

2. 中国科学院机器人与智能制造创新研究院, 辽宁 沈阳 110016)

(1. Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China;

2. Institutes for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences, Shenyang 110016, China)

**摘要:** 介绍了边缘计算的基本概念以及工业互联网智能制造边缘计算的发展现状。通过分析工业互联网、智能制造、边缘计算之间的关系,以及工业互联网智能制造边缘计算行业典型案例,总结了工业互联网智能制造边缘计算的核心问题,并提出了其面对的挑战。

**关键词:** 工业互联网; 智能制造; 边缘计算; 确定性; 可编程性

**Abstract:** In this paper, the basic concept of edge computing and the development status of edge computing in industrial internet intelligent manufacturing are introduced. By analyzing the relationship among industrial internet, intelligent manufacturing and edge computing, and typical cases of industrial internet intelligent manufacturing edge computing industry, the core problems and the challenges of industrial internet intelligent manufacturing edge computing are summarized and put forward.

**Key words:** industrial internet; intelligent manufacturing; edge computing; certainty; programmability

DOI: 10.12142/ZTETJ.201903008

网络出版地址: <http://kns.cnki.net/kcms/detail/34.1228.TN.20190604.1045.004.html>

网络出版日期: 2019-06-04

收稿日期: 2018-12-27

## 1 工业互联网智能制造边缘计算概述

### 1.1 边缘计算及其发展

目前科技处于史无前例的高速发展时代,新的技术层出不穷。近年来,国际著名咨询公司高德纳(Gartner)每年都会发布本年度的新兴技术周期曲线,和下一年度十大战略技术趋势<sup>[1]</sup>,对新兴技术

的发展趋势进行了准确的预测。边缘计算以及边缘智能,与人工智能、深度学习、区块链技术等炙手可热的科技名称并列,连续出现在Gartner公司的咨询报告上。边缘计算已经受到学术界、工业界以及政府部门极大关注,目前学术界发表了很多边缘计算综述,工业界成立了边缘计算产业联盟等多个边缘计算联盟组织,政府部门也发布了一系列边缘计算重大研究计划,人工智能标准化机构也将边缘计算列为人工智能的重要组成部分<sup>[2]</sup>。文

章中,我们主要关注的是中国工业互联网智能制造边缘计算的一些现状与挑战,更大范围的边缘计算的相关情况请见施巍松教授的研究综述<sup>[3-5]</sup>。

边缘计算的发展与云计算、大数据以及物联网技术的发展息息相关,其核心是如何保障网络边缘应用的实时性问题。不同领域应用实时性挑战的来源不同,因此各个领域研究边缘计算的侧重点并不相同,例如人工智能标准化机构将可穿戴式计算列为边缘计算的一个核

基金项目: 国家重点研发计划(2018YFB1700200)、工业和信息化部工业互联网创新发展工程项目“工业互联网边缘计算测试床”

心场景。

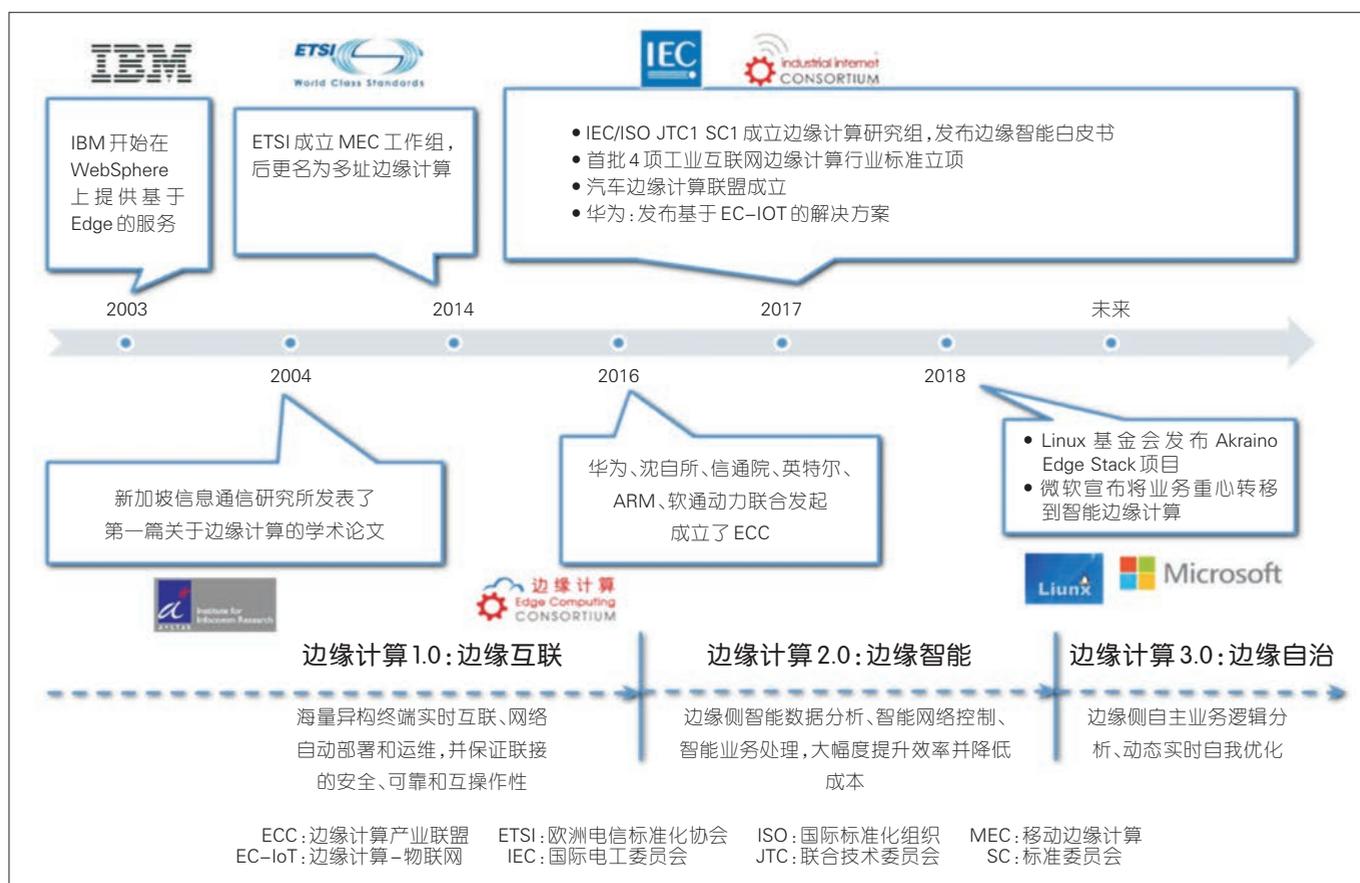
2016 美国韦恩州立大学施巍松教授给出了边缘计算的一般性的定义:“边缘计算是指在网络边缘执行计算的一种新型计算模型,边缘计算操作的对象包括来自于云服务的下行数据和来自于万物互联服务的上行数据,而边缘计算的边缘是指从数据源到云计算中心路径之间的任意计算和网络资源”<sup>[5]</sup>。

虽然近年来受到了广泛的重视,但边缘计算并非一个新的概念。如图 1 所示,2003 年 IBM 已经开始在 WebSphere 上提供基于 Edge 的服务,2004 年新加坡信息通信研究所发表关于边缘计算的学术论文。近年来,随着 5G 和物联网的发

展、智能终端设备不断普及,网络边缘侧数据的爆发式增长推动了边缘计算的发展。2014 年欧洲电信标准化协会(ETSI)成立移动边缘计算(MEC)标准化工作组,同年,AT&T、思科、通用电气(后简称为通用)、IBM 和英特尔成立工业互联网联盟(IIC);2015 年,ARM、思科、戴尔、英特尔、微软和普林斯顿大学建立开放雾(OpenFog)联盟;2016 年,电气和电子工程师协会(IEEE)和国际计算机学会(ACM)共同发起了边缘计算研讨会,中国工业、信息通信业、互联网等领域百余家单位共同发起成立工业互联网产业联盟,同年,由中国科学院沈阳自动化研究所等单位联合倡议发起的边缘计

算产业联盟(ECC)。在学术科研方面,中国自动化学会于 2017 年率先成立了边缘计算专业委员会,并在《IEEE COMMUNICATIONS Surveys &Tutorials》等著名学术期刊上发表了 很多边缘计算综述性文章<sup>[6-9]</sup>。在标准制定方面,国际电信联盟电信标准分局(ITU-T) SG20 发起了“边缘计算需求和能力要求”的国际标准立项,中国通信标准化协会(CCSA)也陆续开展多项边缘计算行业标准立项。

2018 年是边缘计算蓬勃发展的一年:在微软 2018 年度开发者大会上,微软公司发布“Azure IoT Edge”等边缘侧产品,将业务重心从 Windows 操作系统转移到智能边缘



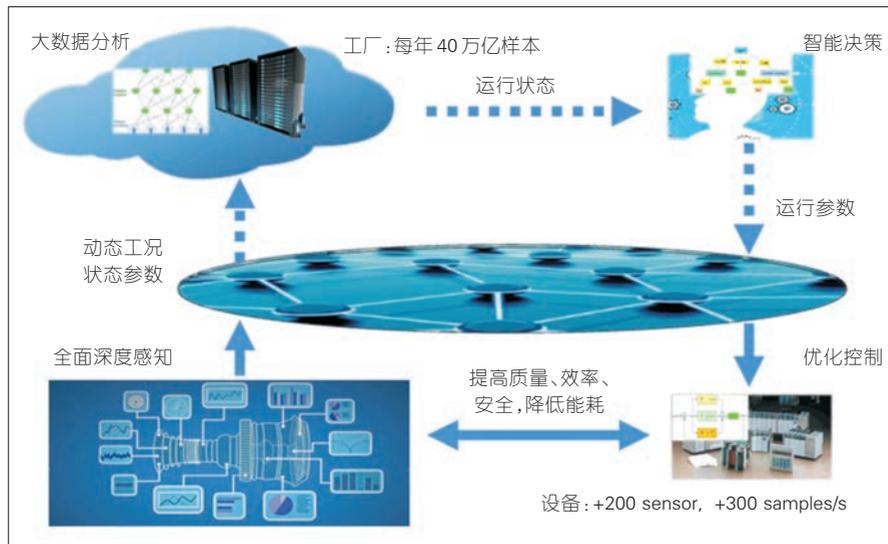
▲ 图 1 边缘计算发展历程

计算方面;亚马逊公司发布“AWS Greengrass”边缘侧软件,将AWS云服务无缝扩展至边缘设备;阿里云宣布2018年将战略投入边缘计算技术领域并推出边缘计算产品Link Edge。

### 1.2 工业互联网智能制造

智能制造是具有信息深度自感知、智慧优化自决策、精准控制自执行等功能的先进制造过程、系统与模式的总称。智能制造可以追溯到1990年4月日本发起“智能制造系统”国际合作研究计划。目前,基于物联网、大数据、云计算等信息技术,智能制造已经贯穿于设计、生产、管理、服务等制造活动的各个环节。工业互联网是指在物联网的基础上,综合应用大数据分析技术和远程控制技术,优化工业设施和机器的运行和维护,通过网络化手段提升工业制造智能化水平。2012年,通用公司首先提出了工业互联网概念,此后工业互联网在世界范围内得到了广泛的发展。目前工业互联网以德国“工业4.0平台”和美国“工业互联网联盟”为典型代表,其分别发布了工业互联网智能制造参考架构RAMI 4.0和IIRA。工业互联网作为中国智能制造发展的重要支撑已经得到了国家的高度认可与充分重视,“十三五”规划、中国制造2025、“互联网+”“深化制造业与互联网融合发展”等国家的重大战略都明确地提出了要大力发展工业互联网。

如图2所示,工业互联网智能化生产、网络化协同、个性化定制和



▲图2 工业互联网智能制造面临的问题

预测性维护等创新应用对计算提出新需求;而工业应用低开销、实时性要求使云计算模式难以适用,而传统的分布式计算方法缺乏高效的网络边缘侧资源管理和任务调度方法,难以应对先进工业生产过程中海量数据的实时分析和系统的实时控制任务。因此,需要发展云与网络边缘侧融合进行数据分析和计算的新模式。

### 1.3 工业互联网智能制造边缘计算

边缘计算是集先进网络技术、大数据、人工智能于一身,横跨通信、计算机、自动控制等多领域的综合性技术。不同领域内的边缘计算技术的侧重点有所不同,工业互联网智能制造对系统的确定性、实时性和安全性有着很高的要求,ECC给出了工业互联网智能制造领域边缘计算的定义:边缘计算是在靠近物或数据源头的网络边缘侧就近提供边缘智能服务,满足行业数字化

在敏捷联接、实时业务、数据优化、应用智能、安全与隐私保护等方面的关键需求<sup>[10]</sup>。工业互联网智能制造边缘计算的体系架构具体如图3所示。

针对工业流数据实时分析、边缘侧智能计算、分布式实时控制等工业互联网智能制造典型应用场景,工业互联网智能制造边缘计算需要解决工业互联网边缘计算场景下计算资源的零散性和有限性、现场网络的异构性对系统确定性、实时性和安全性带来的问题,而工业互联网智能制造边缘计算参考架构的设计、工业互联网智能制造边缘计算的标准化、工业互联网智能制造测试平台和测试方法的构建,是解决上述问题的关键。

## 2 工业互联网智能制造边缘计算现状

以ECC的成立为分界点,工业互联网智能制造边缘计算发展大体可以分为3个阶段:边缘互联、边缘

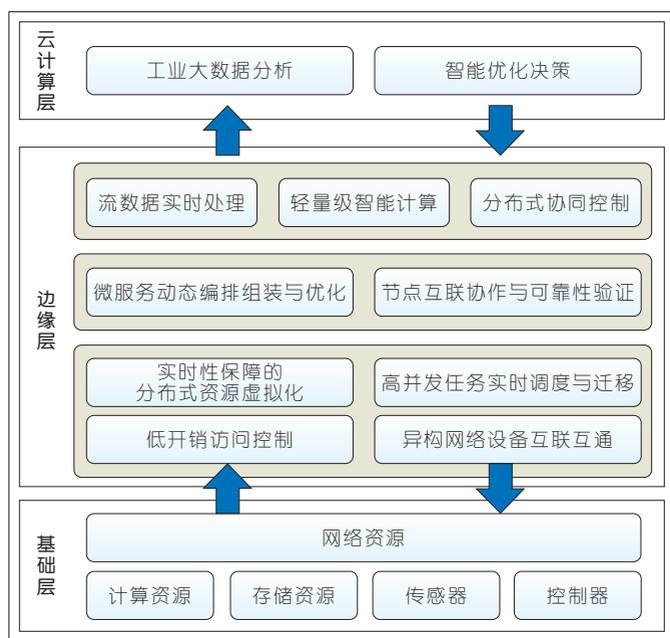


图3 工业互联网智能制造边缘计算的体系架构

智能,以及边缘自治。2016年以前,中国边缘计算的研究处于1.0阶段,主要解决边缘互联问题,即海量异构终端实时互联、网络自动部署和运维等问题。目前,边缘计算处于2.0阶段,即边缘智能阶段,主要解决网络边缘侧智能数据分析、智能网络控制、智能业务处理等问题,从而大幅度提升效率并降低成本。未来,边缘计算将步入3.0阶段,即边缘自治阶段,边缘侧自主业务逻辑分析、动态实时、自我优化将在这一阶段得到实现。

### 2.1 工业互联网边缘计算产业学术组织

#### (1)IIC。

2014年4月,AT&T、思科、通用、IBM和英特尔宣布成立IIC。2017年1月,工业互联网联盟发布工业互联网参考架构IIRA v1.8。目前,工业互联网联盟拥有211个成员单位,涵盖能源、医疗、制造、采

矿、零售、智慧城市、运输等多个领域,并批准了26个工业互联网测试床项目。

#### (2)OpenFog联盟。

2015年11月,ARM、思科、戴尔、英特尔、微软和普林斯顿大学建立OpenFog联盟,目前有60个成员单位。2017年2月,该联盟发布OpenFog参考架构,这是一个旨在支持物联网、5G和人工智能应用的数据密集型需求的通用技术框架。2018年12月18日,OpenFog联盟宣布并入IIC。

#### (3)工业互联网产业联盟。

2016年2月,由工业、信息通信业、互联网等领域百余家单位共同发起成立工业互联网产业联盟。目前,工业互联网产业联盟拥有942个成员单位,发布了“工业互联网平台标准体系框架”等10项白皮书、“工业互联网标准体系框架”等9项技术标准,以及“基于安全可靠云平台的智能工厂验证示范平台”等34

个测试床项目。

#### (4)ECC。

2016年11月,由中国科学院沈阳自动化研究所等联合倡议发起了ECC。目前ECC共拥有成员单位200余家,业务范围包括工业制造、智慧城市、交通、电力能源等一些领域。2018年,ECC发布了“边缘计算与云计算协同白皮书”和“边缘计算参考架构3.0”,同时发布了“面向个性化定制的自适应模块化制造验证平台”等29个边缘计算的测试床项目。

#### (5)中国自动化学会边缘计算专业委员会。

在专业学术团体方面,中国自动化学会率先成立了边缘计算专业委员会。首届中国自动化学会边缘计算专业委员会会议于2017年10月举行,中国科学院沈阳自动化研究所为承办单位。目前,中国自动化学会边缘计算专业委员会委员以项目/课题负责人方式承担了与边缘计算相关的国家重点研发计划、工信部智能制造、工信部工业互联网创新发展工程等一系列重大项目,一定程度上引领了中国边缘计算的发展。

### 2.2 工业互联网边缘计算计算平台

#### (1)通用Predix。

2012年通用提出工业互联网的概念,随后推出Predix。通用Predix分为边缘端、平台端和应用端。在边缘端,Predix提供网关框架Predix Machine以实现数据的采集和连接,支持工业协议解析、数据

采集、多平台协同、本地存储和转发、多种安全策略,以及本地设备通信等。平台端 Predix Cloud 集成了工业大数据处理和分析、Digital Twin 快速建模、工业应用快速开发等方面的能力,以及一系列可以快速实现集成的货架式微服务。在应用端,通用 Predix 采用数据驱动和机理结合的方式,提供设备健康和故障预测、生产效率优化、能耗管理、排程优化等解决方案。2018年10月 Predix 被通用出售。

(2) 西门子 Mindsphere。

2016年4月,西门子推出的基于云的开放式物联网操作系统 MindSphere。MindSphere 向下提供数据采集接入方案 MindConnect,可以直接到达车间级工厂设备,支持开放式通信标准 OPC UA,实现西门子和第三方设备的数据连接;MindSphere 向上为应用软件开发层提供一个开放的架构,用户可以针对不同场景来开发相应的软件。2018年7月,西门子与阿里云签署备忘录,共同推进中国工业物联网发展。

(3) 其他的一些重要的物联网(IoT)平台。

目前各大主流 IT 企业都推出了自己的 IoT 平台,包括 Google Cloud IoT、MS Azure IoT、AWS IoT、IBM Watson IoT 等。例如,Google Cloud IoT 是一个边缘/云服务平台,该平台分为3层:边缘设备、云数据分析、数据应用。Google Cloud IoT 在边缘设备层中集成了一个实时机器学习工具;云数据分析组件包括4个关键功能:Cloud Bigtable、

BigQuery、CloudDataflow 和 Cloud Machine Learning;数据应用层主要用 Cloud Datalab 和 Cloud Studio 解决方案来实现数据的存储和使用。

### 2.3 工业互联网边缘计算标准化

受5G技术发展的推动,ETSI、第3代合作伙伴计划(3GPP)、CCSA、中国联通等通信组织和企业进行了一系列通信领域边缘计算标准化工作;但这些标准化工作主要为了满足5G三大应用场景的需求,即增强移动宽带(eMBB)、高可靠低时延(uRLLC),以及海量机器类通信(mMTC)。

在工业互联网智能制造领域,2017年国际标准化组织(ISO)/国际电工委员会联合技术委员会(IEC JTC1)/标准委员会(SC)41 物联网及其相关技术、分技术委员会成立了边缘技术研究组,并在SC41 范围就边缘计算的定义达成共识,规定边缘计算-物联网(EC-IoT)的参考架构;2018年,SC41 生成现有 EC-IoT 标准的需求分析报告,推荐使用 EC-IoT 技术,为如何实现开放跨行业提供指导,发布了 EC-IoT 项目技术报告;2019年,SC41 将参与边缘计算标准的制定。

2017年,三菱、SAP等联合发布了《垂直边缘智能白皮书》。2018年,工业互联网产业联盟发布了“工业互联网标准体系框架”“工业互联网安全总体要求”等9项技术相关标准。2018年,CCSA 工业互联网特设组(ST8)召开会议,通过“边缘计算总体架构与要求”“边缘计算边缘节点模型与要求”等9项新立项

建议。

### 2.4 工业互联网边缘计算项目设立

工业互联网智能制造边缘计算也受到了国家各部委的高度重视。工信部在2017年和2018年连续设立了一系列智能制造综合标准化与新模式应用项目。2017年,中国科学院沈阳自动化研究所承担的工信部智能制造综合标准化与新模式应用项目“工业互联网应用协议及数据互认标准研究与试验验证”,从工业互联网边缘计算模型、工业互联网数据统一语义模型、工业互联网互联互通信息安全要求等7个方面对工业互联网智能制造边缘计算标准的制定进行了探索。2018年,工信部工业互联网创新发展工程系列项目中,针对工业互联网边缘计算,专门设立了“工业互联网边缘计算测试床”“工业互联网边缘计算基础标准和试验验证”等8个项目。在2018年度科技部国家重点研发计划“网络协同制造和智能工厂”重点专项中,专门针对边缘计算设置了“工业互联网边缘计算节点设计方法与技术”“典型行业装备运行服务平台及智能终端研制”“基于开放架构的云制造关键技术与平台研发”等多个项目。

## 3 工业互联网智能制造边缘计算行业典型案例

### 3.1 中科院沈阳自动化研究所智能制造边缘计算示范线

传统制造系统一般是针对一种

或几种产品而设计的批量化、刚性生产系统,一旦生产需求发生变化,生产线调整周期长,设备维护成本高,难以满足目前智能制造个性化定制的需求。

为了解决上述问题,中科院沈阳自动化研究所搭建了智能制造边缘计算示范系统。如图4所示,该平台由工业云平台、边缘计算数据平台、边缘计算网络、边缘计算网关、现场设备等部分构成。通过引入边缘计算,能够为个性化定制生产带来3方面提升:首先,设备可以灵活替换。通过web互操作接口进行工序重组,实现新设备的即插即

用,实现损坏设备的快速替换,减少50%的人力投入。其次,实现了生产计划灵活调整。通过生产节拍、物料供给方式的自动变化来适应每天多次的计划调整,消除多个型号的混线切单,物料路径切换导致的输入/输出(I/O)配置时间损耗。再次,实现了新工艺/新型号的快速部署。通过Web化的工艺模型的自适应调整,消除新工艺部署带来的可编程逻辑控制器(PLC)(涉及数百个逻辑块、多达十几层层嵌套判断逻辑)重编程、断电启停、数百个过程控制量修改重置的时间,这使得新工艺的部署时间大概缩短了

80%以上。

### 3.2 工业互联网创新发展工程

边缘计算是一个综合性的技术体系,单一技术通常难以反映系统整体性能,因此边缘计算测试床的开发一直是各大边缘计算产业/技术联盟十分重视的内容。2018年,在工信部工业互联网创新发展工程自助下,中科院沈阳自动化研究所承担了我国首批大型边缘计算测试床项目——“工业互联网边缘计算测试床项目”,项目整体架构如图5所示。

该边缘计算测试床项目架构包括3个主要部分:

(1)边缘计算云服务平台,即利用云服务平台实现实验室搭建的生产模拟系统以及汽车装备制造、工程机械、铸造装备加工和光伏装置制造等企业的实际生产系统之间的互联与信息共享。

(2)边缘计算基础测试平台,即为验证边缘计算领域模型,支撑边缘计算关键技术的研发与测试,实现加工、装配、物流工艺过程模拟。

(3)边缘计算垂直行业测试平台,即用于验证边缘计算垂直行业领域模型,支撑边缘计算工业应用APP研发与测试,构建基于企业实际生产系统的行业测试平台,包括:汽车装备制造、工程机械、铸造装备加工和光伏装置制造等企业的4套实际生产系统。

该项目解决工业互联网边缘计算中的问题:边缘计算关键技术研发的试验测试问题、基于边缘计算开发新型工业APP的实验验证问

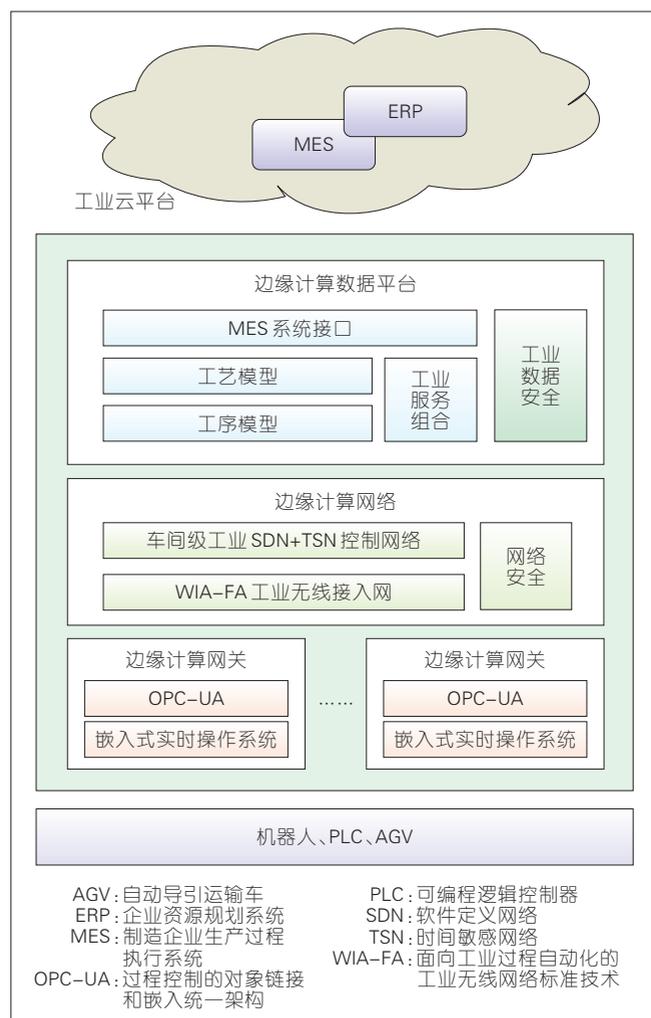
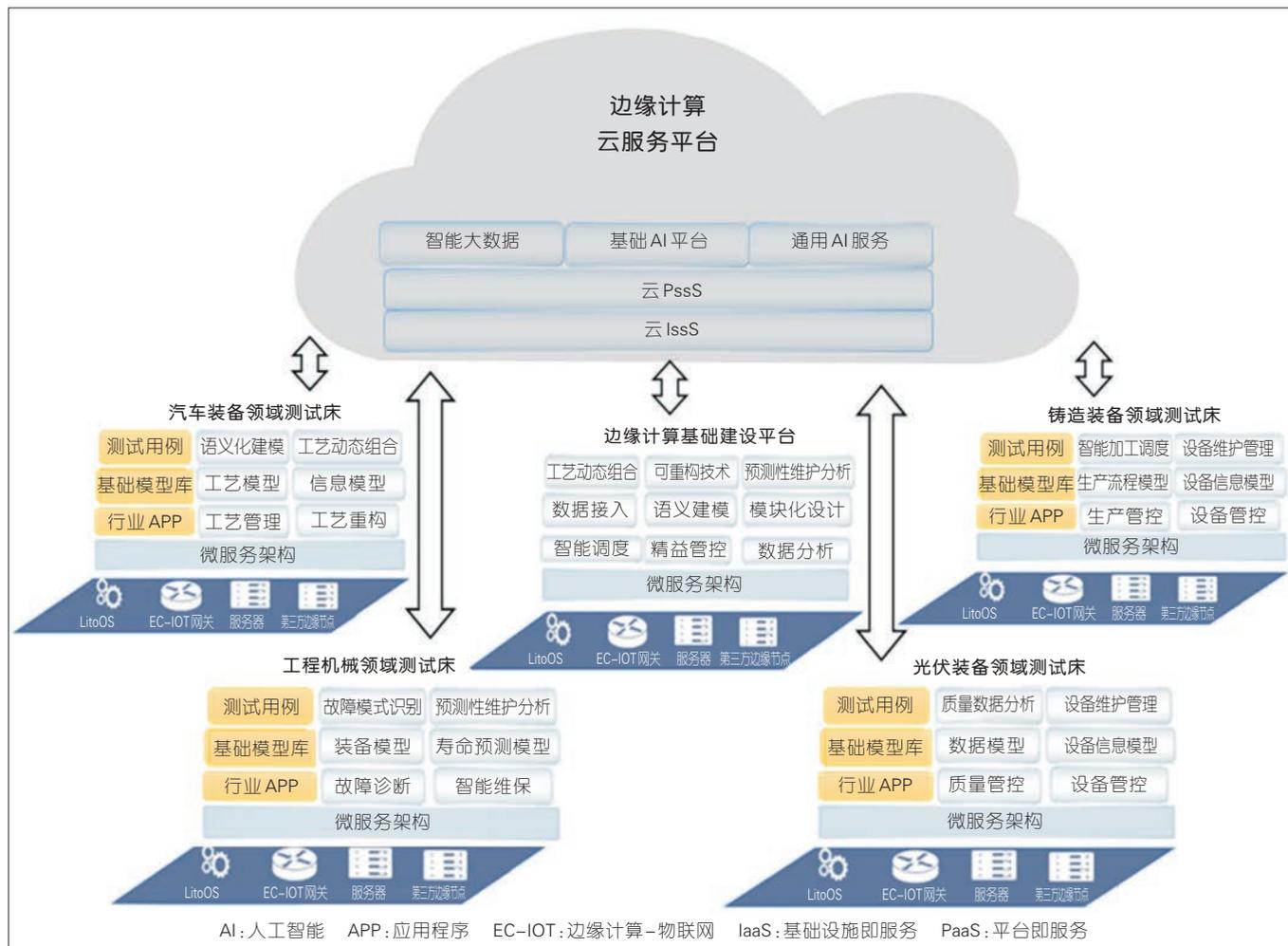


图4 中科院沈阳自动化研究所智能制造边缘计算示范系统整体架构



▲图5 边缘计算测试床项目架构

题,以及基于边缘计算的行业解决方案的多厂商产品互操作问题。

## 4 工业互联网智能制造边缘计算面临的挑战

### 4.1 资源动态条件下高并发分布式混杂系统中确定性保障问题

工业互联网智能制造边缘计算的关键需求是边缘应用的实时性和确定性;但是边缘计算环境存在计算资源分布式、零散化,以及网络动态性和存在测量噪声等特征,并且一般为兼有离散事件和连续变量等

运行机制的混杂系统。同时,边缘计算任务常存在高并发的特点。这些特征导致边缘计算任务时序复杂难以同步,计算结果确定性难以保证。如何针对计算资源的零散性、异构性和动态性、系统的混杂性特点,构建支持分布式混杂系统和高并发任务的边缘计算模型,是工业互联网智能制造边缘计算的挑战。

### 4.2 面向工业互联网中分布式边缘设备的异构行为统一建模与编程问题

工业互联网中边缘计算环境中

的计算、网络和存储等资源的异构性、动态性、分布式和零散化等特性,使得边缘设备资源的动态调整对软件的动态性和可伸缩性要求提升,多种异构资源共存的复杂系统对软件的可组合性与模块化程度需求提高,传统集中式应用程序的开发模式难以满足边缘计算场景的需求。如何解决面向工业互联网中分布式边缘设备的异构行为统一建模与编程问题,实现统一的编程与开发环境,提升编程系统在部署、调试和运行各类应用时的资源利用率,降低部署和维度的难度和时间,是

工业互联网智能制造边缘计算面对的一个挑战。

## 5 结束语

随着物联网技术的发展,网络边缘侧数据呈现爆炸式的增长,云计算模式的缺陷日益突出,而边缘计算为解决工业互联网智能制造过程中数据分析和实时控制提供了有效手段。可以预见:在未来几年内,边缘计算将广泛应用于工业生产之中。把握边缘计算技术发展方向,促进边缘计算技术落地,必将有力地促进工业互联网智能制造的发展。

### 参考文献

[1] Gartner. Gartner Top 10 Strategic Technology Trends for 2019 [EB/OL]. [2019-01-02]. <https://www.gartner.com/smarterwithgartner/gartner-top-10-strategic-technology-trends-for-2019>

[2] 中国电子技术标准化研究院. 人工智能标准化白皮书(2018版)[EB/OL]. [2019-01-02]. <http://www.cesi.ac.cn/images/editor/20180124/20180124135528742.pdf>

[3] 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

[4] 施巍松, 孙辉, 曹杰, 等. 边缘计算: 万物互联时代新型计算模型[J]. 计算机研究与发展, 2017, 54(5): 907-924. DOI:10.7544/issn1000-1239.2017.20160941

[5] SHI W, CAO J, ZHANG Q. Edge Computing: Vision and Challenges [J]. IEEE Internet of Things Journal, 2016(5): 637-646. DOI: 10.1109/JIOT.2016.2579198

[6] MAO Q, HU F, HAO Q. Deep Learning for Intelligent Wireless Networks: A Comprehensive Survey [J]. IEEE Communications Surveys & Tutorials, 2018, 20(4):2595-2621. DOI: 10.1109/COMST.2018.2846401

[7] MOHAMMADI M, AI-FUQAHA A, SOROUR, et al. Deep Learning for IoT Big Data and Streaming Analytics: A Survey [J]. IEEE Communications Surveys & Tutorials, 2018, 20(4):2923-2960. DOI: 10.1109/COMST.2018.2844341

[8] 汪海霞, 赵志峰, 张宏纲. 移动边缘计算中数据缓存和计算迁移的智能优化技术[J]. 中兴通讯技术, 2018, 24(2):19-22. DOI: 10.3969/j.issn.1009-6868.2018.02.004

[9] 王秋宁, 谢人超, 黄韬. 移动边缘计算的移动性管理研究[J]. 中兴通讯技术, 2018, 24(1):37-41. DOI:10.3969/j.issn.1009-6868.2018.01.008

[10] 边缘计算产业联盟. 边缘计算参考架构3.0 [EB/OL]. [2019-01-02]. <http://www.econsortium.org/Uploads/file/20181129/1543443431522383.pdf>

### ← 上接第 36 页

满足应用的时延要求前提下,尽可能降低系统总体能耗,保持各计算节点负载均衡,及时响应紧急计算任务。

### 参考文献

[1] 徐志伟, 曾琛, 朝鲁, 等. 面向控域的体系结构: 一种智能万物互联的体系结构风格[J]. 计算机研究与发展, 2019, (1):90-102

[2] SHI W, CAO J, ZHANG Q, et al. Edge Computing: Vision and Challenges [J]. IEEE Internet of Things Journal, 2016, (3):637-646

[3] XU Z, PENG X, ZHANG L, et al. The Phi-Stack for Smart Web of Things [C]//Proceedings of the Workshop on Smart Internet of Things. USA: ACM. 2017: 1-6

[4] REINSEL D, GANTZ J, .RYDNING J. Data Age 2025: The Evolution of Data to Life-Critical[R/OL]. <https://www.seagate.com/files/www-content/our-story/trends/files/Seagate-WP-DataAge2025-March-2017.pdf>

[5] FIELDING R T, TAYLOR R N. Architectural Styles and the Design of Network-Based Software Architectures [D]. University of California, 2000

[6] CAO J, XU L Y, ABDALLAH R, et al.

EdgeOSH: A Home Operating System for Internet of everything[C]//2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS). USA: IEEE, 2017: 1756-1764. DOI:10.1109/ICDCS.2017.325

[7] DAWSONHAGGERTY S, KRIOUKOV A, TANEJA J, et al. BOSS: Building Operating System Services [EB/OL].[2018-12-22].<https://www.usenix.org/conference/nsdi13/technical-sessions/presentation/dawson-haggerty>

[8] BORMANN C, CASTELLANI A P, SHELBY Z. CoAP: An Application Protocol for Billions of Tiny Internet Nodes [J]. IEEE Internet Computing, 2012, 16(2): 62-67. DOI:10.1109/mic.2012.29

[9] XU Z, CHAO L, PENG X. T-REST: An Open-Enabled Architectural Style for the Internet of Things [J]. IEEE Internet of Things Journal.2019. DOI: 10.1109/JIOT.2018.2875912

[10] HUNKELER U, TRUONG H L, STANFORD-CLARK A. MQTT-S — A Publish/Subscribe Protocol for Wireless Sensor Networks[C]//2008 3rd International Conference on Communication Systems Software and Middleware and Workshops (COMSWARE '08). India: IEEE, 2008. DOI: 10.1109/COMSWA.2008.4554519

[11] SHELLEY Z, BORMANN C. 6LoWPAN: The Wireless Embedded Internet [M]. USA: Wiley, 2009

### 作者简介



**宋纯贺**, 中国科学院沈阳自动化研究所研究员, 中国科学院百人计划入选者, 中国自动化学会边缘计算专业委员会秘书长; 研究方向为工业互联网边缘计算; 发表论文 50 余篇。



**曾鹏**, 中国科学院沈阳自动化研究所所长助理、研究员, 中国科学院网络化控制系统重点实验室主任, 中国自动化学会边缘计算专业委员会主任, 国家“万人计划”创新领军人才; 研究方向为工业互联网; 获得国家科技进步一等奖等省部级科技奖励 6 项; 发表论文章百余篇。



**于海斌**, 中国科学院沈阳自动化研究所所长、研究员, 国家杰出青年基金获得者, 国家“万人计划”创新领军人才, ISA Fellow, 中国自动化学会副理事长, 边缘计算产业联盟理事长; 研究方向为工业互联网; 获国家科技进步二等奖 4 项、技术发明二等奖 1 项; 发表论文章百余篇。

### 作者简介



**彭晓晖**, 日本东京大学博士毕业, 中国科学院计算技术研究所副研究员、硕导; 现主要从事物端计算系统的编程方法、应用协议与系统架构研究, 研究方向为分布式系统、边缘计算等; 主持和参与了数项国家、部委级研究课题; 发表论文章 20 篇。



**徐志伟**, 中国科学院计算技术研究所研究员、博士生导师、学术委员会主任, 中国科学院大学教授, CCF 会士, 曾担任国家自然科学基金委信息科学部专家咨询委员会委员, 《IEEE TC》《IEEE TSC》等多个国际著名期刊的编委等; 主要研究领域为高性能计算机、分布式计算、网格计算与系统软件, 近年来主要从事边缘计算、海云计算系统的研究, 并提出了物端计算方向; 曾获国家杰出青年科学基金、中国科学院科技进步一等奖、国家科学进步二等奖、中国科学院优秀青年科学家奖、中国科学院百人计划资助、中国科学院杰出科技成就团队奖等, 2007 年获中国计算机学会“王选奖”; 发表论文章 100 余篇、专著 3 部。

# 面向无人驾驶的边缘 高精地图服务

## Edge High-Precise Map Services for Autonomous Driving

唐洁/TANG Jie<sup>1</sup>, 刘少山/LIU Shaoshan<sup>2</sup>

(1. 华南理工大学, 广东 广州 510006;

2. 深圳普思英察科技有限公司, 广东 深圳 518000)

(1. South China University of Technology, Guangzhou 510006, China;

2. Shenzhen PerceptIn Technology Co., Ltd., Shenzhen 518000, China)



**摘要:** 高精度地图作为无人驾驶应用中辅助驾驶的重要手段, 在高精度定位、辅助环境感知、控制决策等方面发挥着重要作用。边缘计算将计算、存储、共享能力从云端延伸到网络边缘, 使用“业务应用在边缘, 综合管理在云端”的模式, 非常适合于部署更新频率高、实时服务延迟低、覆盖面积广的高精地图服务。结合众包的边缘地图服务已经成为高精地图更新的主要方式, 有着重要的应用前景。

**关键词:** 高精地图; 无人驾驶; 边缘计算; 众包

**Abstract:** High-precision map plays an important role in autonomous driving, and it is the important foundation for high-precision positioning, environmental perception and control decision. Edge computing extends computing, storage and data sharing from cloud to network edge thus it is very suitable for deploying high-precision map services. With the mode of "deployment in edge, management in cloud", edge high-precision map service can provide high update frequency, low real-time service delay and wide coverage area. Combined with crowdsourcing, edge high-precision map has become the applicable and promising solution for low-cost and high-precision map updating.

**Key words:** high-precision map; autonomous driving; edge computing; crowdsourcing

DOI: 10.12142/ZTETJ.201903009

网络出版地址: <http://kns.cnki.net/kcms/detail/34.1228.TN.20190605.1604.004.html>

网络出版日期: 2019-06-05

收稿日期: 2018-12-15

无人驾驶作为目前人工智能行业最受关注的应用场景之一, 担当着革新汽车行业甚至是交通运输业未来的重要使命。无人驾驶能够真正地解放人类双手, 提高行车安全, 通过更普及的运力共享, 在缓解交通拥堵的同时可大大地减少对环境的污染。随着无人车的普及, 无人驾驶将成为未来智慧公共出行的主要方式, 是未来智慧城市的重要联结之一<sup>[1]</sup>。

在无人驾驶应用中, 高精地图

是其必不可少的实现基础。高精地图是对物理世界路况的精准还原, 通过道路信息的高精度承载, 利用超视距信息, 和其他车载传感器形成互补, 打破车身传感的局限性, 实现感知的无限延伸。以底层的高精度地图数据为基础, 在此之上叠加动态交通数据, 通过高速通信完成交通信息的实时更新以及驾驶预警推送, 为无人驾驶行车决策提供强有力的指导<sup>[2-3]</sup>。

相比于传统的电子地图, 高精

地图具有数据高精度、信息高维度以及高实时性的特点。随着感知范围的延伸和传感精度的提高, 高精度地图有能力构建更精确的定位、更广范围的环境感知、更完备的交通信息, 从而为无人驾驶提供感知、定位、决策等多种支持<sup>[4-5]</sup>。高精度地图不仅仅包含对道路静态元素 10~20 cm 厘米精度的三维表示, 如车道线、曲率、坡度和路侧物体等, 还包括了驾驶环境中各种动态信息, 如车道限速、车道关闭、道路

坑洼、交通事故等。此外,高精地图还发展出个性化驾驶支持,包括各种驾驶行为建议,如最佳加速点及刹车点、最佳过弯速度等,以提高无人驾驶的舒适度。

目前,全球图商以及无人驾驶车商,如百度、Google、高德、HERE、TomTom等已经组建了自有专业地图采集车队,通过配置有摄像头和激光雷达等设备的高精度地图采集车扫描获得街景图像数据和3D激光点云数据,经过后台的自动化建图流程,结合人工纠错与标注,最终形成多层次地图数据叠加的高精度地图,并进行发布。然而,自建专业采集车队极其昂贵、维护成本开销大且覆盖与更新面积有限,难以实现高精度地图生产的实时更新或修复自愈。地图众包是高精度地图生产与服务提供的新方向。利用(半)社会车辆在行驶过程中完成传感数据采集,通过边缘计算节点的数据清理、聚合和压缩等优化手段,可以将抽取过的关键感知数据推送至云端,在利用云端的强大算力使用多源数据完成对地图数据的更新,最

后再把增量更新与动态实时交通存储于边缘缓存,根据车辆的驾驶场景,完成对高精地图数据的推送与预取。由此可见,边缘高精地图服务是自动驾驶在边缘计算场景下的一大典型应用,在减少成本开销的同时,实现覆盖更为广泛的、更高频的、更实时的地图数据服务。

通过众包机制,利用多车传感在边缘计算端与云端进行协同,实现高精度地图的构建与实时更新;并通过边缘缓存向车辆实时发布高精度地图的动态层数据与静态更新,以辅助车辆的无人驾驶。这一未来无人驾驶典型应用需要解决3大方面的问题:(1)如何以边缘节点为中心,根据边缘智能感知的车辆动态分布进行感知任务分配,并对汇集的群智信息的时空有效性与数据质量进行评估和控制;(2)多车辆传感数据如何在边缘节点自动化地进行过滤、聚合、协商,以得出对交通态势的动态描述,抽取对静态更新的一致性感知结果,其中包括如何通过有序协作对感知数据进行动态融合,从而提高感知准确性、全

面性,减少信息冗余度;(3)根据感知数据的时空特征和车辆的分布规律,如何利用边缘环境中具有时空约束的服务数据缓存与分发,把确认后的地图更新与实时交通状况传播到其他相关车辆中,为无人驾驶服务提供细粒度、准确、高时效的数据基础。

## 1 高精地图

高精地图是高精度、高维度、高实时的地图数据综合。在数据精度上,相比电子地图米级的数据误差,高精地图的相对精度误差不超过20 cm。相比电子地图中简单的道路模型,高精地图包含了道路标志、车道边界、车道坡度、弯道曲率等多维度的丰富信息。同时,通过与车联网以及车到万物(V2X)等技术的结合,高精地图还将更新实时交通信息以及天气状况等。所有这些路况信息是高精地图辅助实现L3~L5级无人驾驶的基础。如图1所示,高精度地图在无人驾驶中发挥着重要作用。通过比对车载传感环境信息与高精地图信息,可以得到

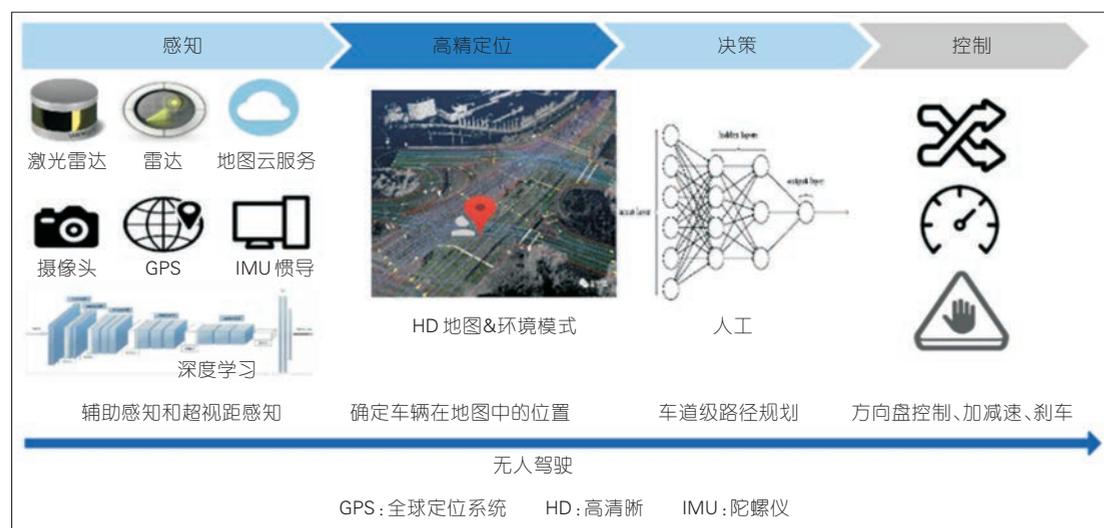


图1  
高精度地图在无人驾驶中的作用

车辆在地图中的厘米级精确位置,实现车辆的高精度定位<sup>[6]</sup>。利用先验知识,高精地图信息可以辅助环境感知,从而降低车载环境感知的难度<sup>[7]</sup>。高精度地图现已精确到车道模型,可以预知前方多种道路信息,实现提前减速和避让,完成更有效的局部驾驶规划。通过5G等通信手段,结合V2X交互支持,可以实现超距离路况感知与预警,完成更智能、更实时全局驾驶规划。

### 1.1 高精地图数据组织

高精度地图作为自动驾驶的必要支撑,在维持底层车道数据精确性的同时,必须具备进行动态路况信息实时更新的能力,并发展出基于车主不同驾驶习惯的个性化驾驶支撑能力。因此,高精地图需分为2个层级:最底层是静态高精地图层,需提前进行装载;上层是动态高精地图层,在行车过程中不断更新。

为了提升存储效率和可读性,静态高精地图在存储时又继续划分为矢量层和特征层。特征地图是对路面信息的准确刻画,主要驱动于高精度定位的需求,能够很好地完成地图道路匹配与定位。矢量地图是在特征地图基础之上进一步的抽象、处理和标注。它的容量更小,并能够通过其中的路网信息完成点到点的精确路径规划,这是高精度地图使能的一大途径。如图2所示,矢量层包含车道模型、道路部件、道路属性数据。这些语义元素被简化和抽取出来,补充到几何构建的道路结构中,形成新的矢量地图数据。其中,车道模型中包含了车道

图2  
高精地图静态层次内容



线、车道中心线、车道属性变化等信息,可以辅助车辆完成横向定位,并且执行交通规则,比如指导车辆在虚线区域内进行并线,在车道分离点前完成变道。通过对比车载传感数据与交通标志牌等道路部件信息,可以修正车辆纵向定位和航向。即便在没有检测到任何道路特征的情况下,也可以通过高精地图的航位推算进行短时间的推位置推算。车道模型中曲率、坡度、航向、横坡等数学参数,可以决策车辆准确的转向、制动、爬坡等行为。这部分数据因为道路翻新和维护出现频繁变化,需要有效地利用边缘群智感知进行更新。

动态高精地图层构建在静态层之上,增加了道路拥堵情况、施工情况、交通事故、交通管制情况、天气情况等动态交通信息。这些动态要素通过5G等通信手段在车车之间、车与边缘节点之间、多边缘节点之间进行收集、处理与发布,将及时地反映在高精地图上以辅助决策确保无人车行驶安全。

### 1.2 高精地图传感器类型

高精地图的生产依赖于多个传感器数据的融合。高精地图数据采

集时使用的传感器种类不一定与无人驾驶时使用的传感器相同。为保证L4或者L5无人驾驶的要求,高精地图生产需要的传感器主要包括图3中的几种类型。

- 光学摄像头。通过车载摄像头,可以捕捉车身周围交通环境的静态信息,通过对图片中关键交通标志、路面关键信息的提取,来完成对地图的初步绘制。目前,基于图像的深度学习技术发展远超过基于3D点云的深度学习,对于各种地物的提取具有较高的精确率和较低的召回率,可以非常清晰地提取道路边线和交通标识等信息。这些都是摄像头作为高精地图感知手段的重要优势。然而,摄像头传感要求外部环境光线充足,因此采集工作只能在白天进行<sup>[8]</sup>。

- 陀螺仪(IMU)。IMU配备有6轴运动处理组件,包含了3轴加速度和3轴陀螺仪,分别检测在上下、左右、前后这3个方向上的加速度和角速度信息,以此解算出物体的姿态,提供短时内较为准确的定位。但是,从加速度推算出运动距离需要经过2次积分,随着时间漂移所产生的误差将不断增大。因此,仅依赖IMU无法完成长时间的



▲图3 高精地图主要传感器类型

车辆位置精确预测,高精地图的生产往往需要精度较高的IMU,所使用的IMU价格一般在数万到几十万之间<sup>[9]</sup>。

- 轮测距器。车辆的前轮通常配备了轮测距器,分别记录左轮与右轮的总转数。通过分析每个时间段内左右轮的转数、向左右转动的角度等,可以推算出车辆前进的距离和位置。但是,不同的地面材质(如冰面与水泥地)上的转数对距离转换存在偏差。所以单靠轮测距器并不能精准预测无人车位置,但对于高精地图的制作不是必须的传感器。

- 全球定位系统(GPS)。GPS作为使用最广的定位系统,在无人驾驶定位和高精地图制作中都发挥着举足轻重的作用。GPS使用4颗或更多卫星的位置(储存在星历中)计算出地面接收器与每颗卫星之间的距离,然后利用三维空间的三边测量法推算出车辆的位置。民用GPS的单点定位精度一般在米级,差分GPS通过增加一个参考基站的方法可提高定位精度至厘米级<sup>[10]</sup>。但是在复杂环境中,一旦出现信号阻挡的情况,GPS多路径反射(Multi-Path)的问题会导致几十厘米甚至几米的误差。同时GPS的定位频率较低,最高仅为10 Hz,目前

通常使用GPS和IMU连接组成惯性导航系统,利用低频率的GPS数据校准高频但易漂移误差的IMU数据,进而完成相对高频率、高精度的融合数据。

- 激光雷达(LiDAR)。激光雷达通过向目标物体发射一束激光,然后根据接收-反射的时间间隔确定目标物体的实际距离。根据距离以及激光发射的角度和感光测量数据,通过简单的几何变化即可推导出物体的三维空间位置信息和光强度信息。使用激光雷达能够快速获得道路及周边环境的三维点云数据和光强度信息,完成对检测到物体的初步判断。激光雷达的工作条件基本不受时间影响,白天黑夜都可以工作。但是,激光雷达对环境要求较高,在大雨、浓烟、浓雾、大雪等恶劣天气里都会受到很大的影响。同时,激光雷达并不能识别和理解交通标志牌、信号灯等内容,需要借助光学摄像头来捕捉并提取这

部分信息。最重要的是,激光雷达价格昂贵,尤其是高线束激光雷达的量产率比较低,价格居高不下,导致激光雷达的部署成本过高,难以落地。

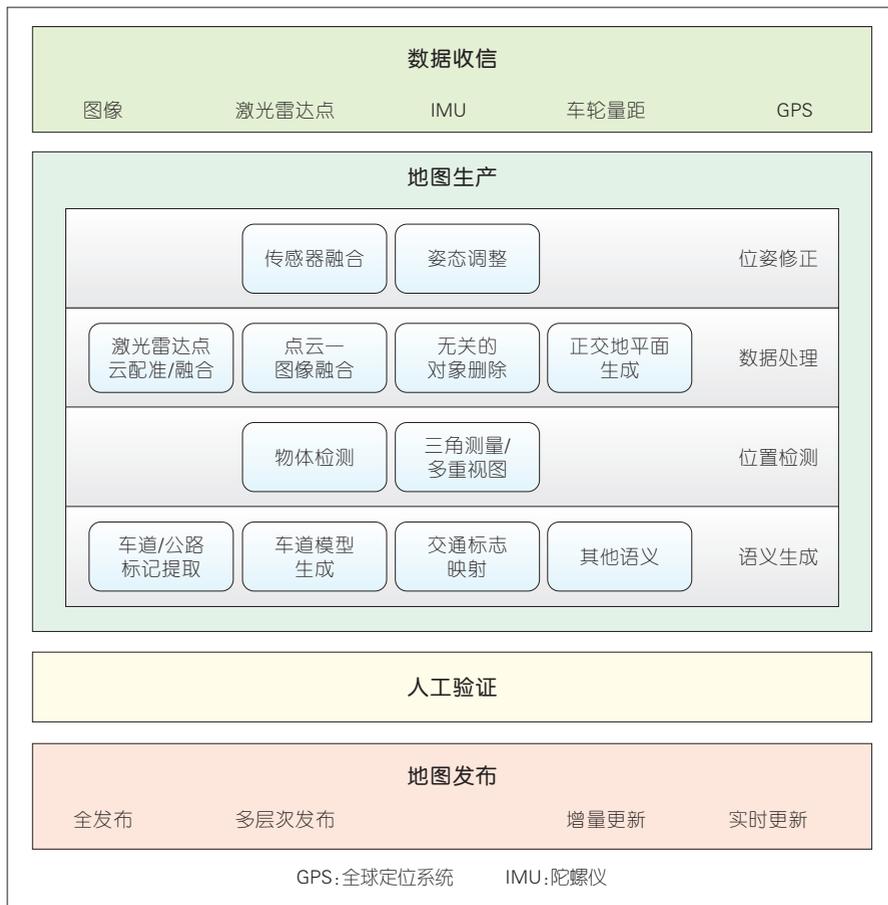
- 多传感器融合。如表1所示,单一传感器很难全天候工作,并且在实际应用中不同传感器有着不同的特性。比如,GPS传感器能够直接测量出每个时间点车辆的全局位姿,但是这些直接测量结果通常不会很精确。激光雷达和摄像头的每次采样测量结果均是在局部坐标系,但是两者包含了丰富且准确的局部测量信息,很多相对的位姿信息可以被抽取出来丰富地图的语义。因此,在制图中均采用多传感器融合的方法。多传感器数据融合就是把所有的车载传感器数据标定到统一的车辆坐标系(如x轴向前,y轴向左,z轴向上)上,并尽可能利用各种离散时空数据,计算出车辆每个时间点在全局坐标系下的三维位姿。一旦有了全局车辆的位姿信息,就可以完成各类传感数据局部测量结果融合。

## 2 高精地图的制作

如图4所示,高精度地图的构建由数据采集、地图生产、人工验

▼表1 各类传感器性能分析

|         | 相机 | 激光雷达 | 毫米波雷达 |
|---------|----|------|-------|
| 目标检测    | 一般 | 优秀   | 优秀    |
| 识别精度    | 一般 | 优秀   | 一般    |
| 目标分类    | 优秀 | 一般   | 差     |
| 视野      | 优秀 | 优秀   | 一般    |
| 不良光线下表现 | 一般 | 优秀   | 优秀    |
| 不良天气下表现 | 差  | 一般   | 优秀    |



▲图4 高精地图生产流程

证、地图更新发布4个过程组成。

### 2.1 数据采集

目前主流的高精度地图数据采集可以分成3大阵营。其一,自由专业采集体系是图商利用自建的高精专业采集车进行上路采集。其二,行业采集生态是利用如物流车、出租车等车辆上的GPS、摄像头等回传实时的道路轨迹和路况信息,这也是现在导航地图中使用的交通动态信息采集方法之一。其三,众包模式。在自动驾驶时代,普及后的每辆无人驾驶车上的激光雷达、摄像头等传感器无时无刻都在采集道路信息。这些传感信息可以被用

来完成对高精地图的更新发现与实施。边缘地图更新就是利用众包在提供边缘地图服务的同时,收集服务车辆的传感数据,利用边缘智能完成对地图多层次数据的迭代。

第一阵营的代表是谷歌、百度等,他们拥有专业的地图测绘车车队,可完成封闭的集中式制图。目前,谷歌汽车已经完成了累积超过1 931 212.8 km的无人驾驶高精度地图测绘<sup>[1]</sup>。然而专业的地图采集车造价非常昂贵,成本达到800万元,相对精度在10 cm之内,这显然是单纯依靠摄像头提供视觉方案的众包模式所无法实现的。第三阵营的代表是Mobileye。Mobileye借助

不同品牌大量级的车载摄像头获取数据来源,并针对自动驾驶情景,将重点放在路上的各种导流标志、方向标识、信号灯等,依靠这些建立的路标,从微观上在行驶过程中为车辆提供指引。

未来在无人驾驶发展和应用中,3种模式将长期共存。地图采集车高昂的造价限制了地图数据大范围、高频率的采集更新,因此这部分数据将作为高精地图基础数据。在此之上,物流车等半社会化商用车和私用车将以众包模式实现对高精度地图数据的补充和更新;而边缘地图服务正是基于此种应用需求,在网络侧完成对众包车辆对道路多源描述的整合。

### 2.2 地图生产

地图生产主要涉及位姿修正、数据预处理、位置检测和语义生成4个环节<sup>[2]</sup>。这一过程将会利用激光点云识别技术、深度学习图像识别技术以及大数据的处理能力等实现多种传感数据的自动融合、识别、语义标注等。一般来说,采集的设备越精密,采集的数据越完整,需要算法修正的不确定性就越大,像Google、HERE等公司已有低分辨率的母图做基础,只需要叠加更多立体图层和语义层。相反,如果采集的数据误差越大,就越需要依赖算法弥补数据的缺陷,对算法要求更高。在边缘地图服务中,来自不同车辆的传感信息本身就存在着各自的误差。同时,不同的传感数据描述通常是对同一事件的抽象,这就对传感数据的边缘抽象、融合、处理

水平提出了较高要求。

### 2.3 人工验证

人工验证这一环节由人工完成。自动化验证的数据还不能达到百分百的准确,需要人工再进行最后一步的确认和完善。

### 2.4 地图更新发布

地图更新发布主要针对静态地图层的道路的修改和动态地图层的突发路况、交通事故等。最终形成的地图更新可在云平台决策之后经由边缘节点和车间通信进行内容的分发。

## 3 边缘场景下的高精地图服务

由于高精度地图对数据更新提出了很高的要求,实时更新和实时同步是高精度地图应用过程中绕不开的2大问题。为了解决这2点,云平台是高精地图所不可或缺的。但是,云平台在高精地图中的直接应用面临2个难点:(1)实时更新、数据同步的困难;(2)云平台制图能力的有限性,包括但不限于数据收集、运算、交互、分发等。因此,高精地图生产与服务更需要从云-边缘-端的角度推进,在分散云中心计算压力的同时,还要强化云-边缘-端之间的联系以及网络侧本身的计算、收集与发布能力。

因此,从高精地图的产品形态和服务方式的角度,通过边缘计算服务对高精地图数据进行实时更新与分发是一种可行的方式。根据边缘计算产业联盟的定义,边缘计算

是在靠近设备或数据源头的网络边缘侧,融合网络、计算、存储、应用核心能力的开放平台,就近提供边缘智能服务,以满足行业数字化在敏捷联接、实时业务、数据优化、应用智能、安全与隐私保护等方面的关键需求<sup>[13-14]</sup>。基于边缘计算的高精地图服务包括地图生产和地图发布2部分内容。边缘地图生产服务通过实时收集各车的行驶数据来增强交通时态的收集密度,扩充道路情况信息的感知范围,并通过感知数据的预处理实现有效的内容提取和无关语义去除等。边缘地图发布服务通过边缘缓存能力进行地图数据发布,将大大地缓解数据更新的缓存开销和到达延迟,实现更贴近用户的数据服务和行车预警。

高精地图对于地图数据处理有着特殊的要求:一是低时延,在车辆高速运动过程中,要实现动态地图中的碰撞预警功能,通信时延应当在4 ms以内;二是高可靠性,高精度地图服务于无人驾驶,相较于普通数据处理,高精地图的传感数据处理需要更高的可靠性。与此同时,车辆的高速运动以及可预见的传感数据量爆发,对于时延和可靠性的要求也将越来越高。边缘计算在局域内即可实现对实时传感数据的聚集、分析与抽取,一方面将分析所得结果以极低延迟(通常是毫秒类)传送给区域内的其他车辆,一方面将抽取后的信息推送至云端,以便地图云完成对更新的决策。通过利用边缘计算的位置特征,地图数据就可实现就近存储,因此可有效降低时延,非常适合于动态高精地

图中防碰撞、事故警告等时延标准要求极高的业务类型。同时,边缘计算能够精确地实时感知车辆移动,提高通信的时效与安全。在此方面,德国已经研发了数字高速公路试验台来提供交通预警服务,该试验台用于在长期演进(LTE)环境下在同一区域内进行车辆预警消息的发布<sup>[15]</sup>。相比集中式高精地图服务,边缘高精地图服务拥有以下的特性:

- 低时延。边缘高精地图服务利用V2X等近距离通信技术来提升行车安全与交通效率。目前边缘节点有效范围内的主要通信方式是专用短程通信(DSRC)、蜂窝通信(LTE-V2X)、5G-V2X。5G网络延迟可以达到毫秒级,峰值速率可达10~20 Gbit/s,连接密度可以达到100万/千米<sup>2</sup>,可保障大规模行车场景下的4 ms碰撞预警时间。这使得驾驶反馈更加迅速,改善了用户安全与用户体验。

- 低负载。由于边缘地图服务更靠近车辆,更适合捕获和分析传感地图数据中的关键信息,可就近在本地进行数据聚合处理,将校验、简化、抽取过的传感信息上传至云端,这将极大地降低核心网的数据传输压力,大大减少网络堵塞。

- 基于位置的内容感知。边缘地图服务节点能实时获取车辆位置相关数据,并使用获得的时空信息对边缘缓存内容进行自适应感知和调整,根据当前行车环境下的车辆分布与内容流行度分布,实现了快速的基于位置的地图应用部署,极大地提高无人驾驶用户的地图服

务使用体验。

- 低成本。边缘高精地图服务不需要专用的地图采集车,通过众包的方式实现对高精地图生成与维护的任务分解,制图成本将大大地降低。

#### 4 边缘高精地图生产

利用感知数据在时空维度存在的潜在模式或相关性,在云地图中心的协调下对边缘节点内的地图生产过程进行优化,根据区域地图特性和实时车流分配感知任务,从而实现边缘资源在地图生产上的动态配置,提高路况感知的效率。因此,大规模地图数据感知的众包任务集确定、传感数据处理是边缘地图生产服务中亟待解决的问题。这一问题主要面临的挑战包括:

(1)如何根据感知数据的时空相关性及车辆的动态分布选择兼具代表性和可行性的感知任务集,即如何进行感知任务分配。

(2)如何对大规模的车辆感知数据进行汇聚和分析,控制众包数据的采集维度、采集频度、采集品类等,都将对最终多维度地图数据的生成产生重大影响。

(3)高精地图相对普通地图拓扑结构更复杂,信息维度更高;相对电子地图,高精度地图精确到厘米,精确度提升2个单位量级。如何在边缘节点完成高精地图裸数据的信息去噪、抽取、处理与标注等操作,都对边缘节点数据利用有限算力进行数据(预)处理自动化提出了更高的要求。

(4)由于参与地图感知任务的

大都为非专业的(半)社会车辆,其行驶轨迹取决于实际需求,因此感知资源将出现时空分布不均匀的现象,影响整体感知质量。为了在边缘节点通过挖掘感知数据的潜在特征和车辆的分布规律进行感知数据的质量评估和控制,就必须考虑如何在降低信息冗余度的同时提高信息完整性与准确性。

#### 5 边缘高精地图内容分发

高精度地图的数据分发需求可达3~4 GB/km,在全球移动数据流量复合年均增长率为47%的场景下,移动通信网络的数据传输延迟将严重限制无人驾驶应用的普及。无人驾驶车辆虽然可以从云端实时获取相关位置的高精度地图数据,但这种数据获取具有相当严格的时空约束。边缘计算基于位置感知提供服务,相近时空内的无人驾驶车辆对于地图数据的服务需求存在广泛的相似性。路径重合的行驶车辆具有相同的地图数据需求,这些时间上离散的地图数据可以通过“边缘缓存+车辆间内容分享”的方式进行地图数据的分发,并且车辆可以通过边缘缓存预取未来途经路段的地图数据,提前完成超距感知和长距离规划;因此,边缘智能协调下的边缘缓存及车间地图内容分享将成为实现高效的地图数据服务的重要途径。基于行驶场景和车辆移动的边缘地图数据缓存和分发方法,通过学习车辆的地图服务请求规律和移动规律智能地进行地图数据缓存,并根据车辆的实时地图数据请求分布、情景预测以及物理位置分

布协调多层次的地图数据分发。为实现此目标,需要解决以下几个重要问题:

(1)车辆的地图数据内容需求,在边缘缓存侧完成数据的层次化分发与预取。

(2)根据车辆出行情景分布的时空变化动态地调整边缘缓存内容,并结合数据编码和数据压缩等技术进一步提高边缘缓存的数据存储能力。

(3)车辆之间如何通过直接通信进行地图内容数据的再分发,最终满足所有车辆的数据需求,以保证服务质量。

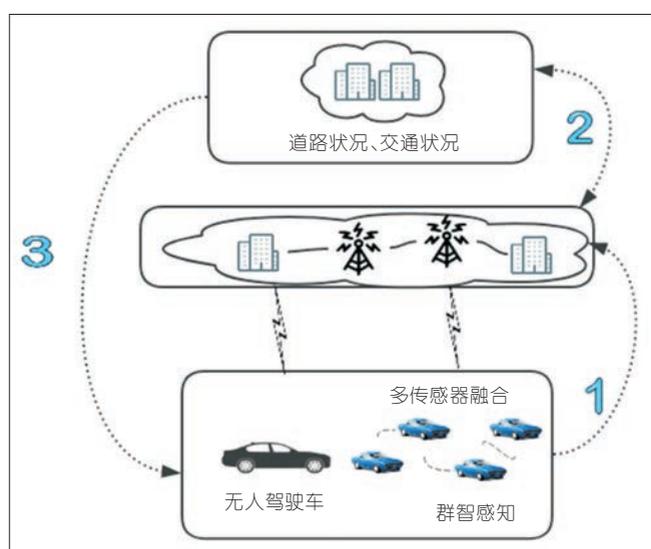
#### 6 参考框架

图5给出了博世高精地图系统示意图。博世地图系统会根据地图数据采集任务进行智能调度,向摄像头和毫米波雷达设备下发采集任务,并通过云端控制完成采集成果自动上传。上传到博世道路特征云的数据,需进行自动化处理,自动地识别其中的有效信息(如道路指示牌、限速标志、车道线标志、交通信号灯等)。所形成的特征内容与地图合作商(如高德、四维)提供的高精地图图层(矢量数据层)进行叠加,形成内容完整的静态高精地图,再由主车厂商在车身上进行前装发布。这里利用了众包机制完成了对特征图层的构建,但是这样的事前行为无法对发生的地图变化进行更新,同时也无法利用边缘服务实现对无人驾驶的数据支持。

为支持安全高效的智能驾驶服务,我们给出了如图6所示的边缘



▲图5 博世高精地图系统示意图



▲图6 边缘高精地图服务参考框架

高精地图服务参考框架。第1部分,利用多车群智协作来实现感知数据的采集与动态地图信息的有效传播,快速的交通信息施放构成“超视距感知”,可辅助车辆进行实时驾驶决策和短期规划。第2部分,大规模的车辆感知数据在边缘端进行汇聚、分析、感知质量判断与控制,完成感知任务调度以实现全区域内的传感能力覆盖,并进行边缘地图自动化制图。此阶段的目的是构建多维度的地图更新要素,并交由云端决策。借助边缘地图服务的高效数据缓存与分发,利用增量更新技术,实现(半)小时级的增量发布。

第3部分,云端汇总全局环境的认知结果和更新要素,通过与云地图数据库进行自动差分比对,可快速定位需要更新的内容,完成对地图内容的迭代。迭代后的地图版本或以前装的形式,或以边缘缓存增量发布的形式装载至无人驾驶车辆,以支持车辆在全局感知下进行中长期规划。

## 7 相关工作

边缘高精地图服务是无人驾驶产生之后,在高精地图需求之下驱动生成的新方向。但是,前人已经在边缘计算协同、边缘缓存和交通

数据感知等领域积累了大量的研究成果。

### 7.1 边缘计算

欧洲电信标准化协会(ETSI)在2015年发布边缘计算白皮书,首次将边缘计算定义为:“通过在移动网边缘部署通用服务器,从而为无线接入网提供IT服务环境和云计算能力”<sup>[16]</sup>,并完成对边缘计算平台架构、计算需求、应用程序标准接口的定义。第3代合作伙伴计划(3GPP)将边缘计算列入未来5G时代的关键技术,在3GPP的标准化架构中将边缘计算的需求作为重要设计元素,给出了边缘计算的业务连续方案和流量疏导方案<sup>[17]</sup>。

在边缘节点上如何进行计算任务的高效卸载和计算协同,是计算在边缘下沉时亟待解决的重要问题。通过边缘-云之间的计算协同,在网络边缘对部分紧急任务进行卸载和计算,对于紧急程度较低的、算力要求高的任务再上传到云数据中心进行处理<sup>[18-19]</sup>,可降低核心网络的计算能耗,提供对关键应用的响应能力。文献[20-21]中,作者列举了边缘计算的具体应用场景,并讨论了在计算卸载过程中可能存在的问题。文献[22]中,作者以实现通信、计算、存储3类基础资源的有效共享为目的,提出了5G移动通信架构下多级计算协同的边缘计算模型。文献[23]中,作者对边缘计算下的任务卸载问题进行了建模,通过高可靠的任务卸载算法,以多用户下的边缘资源分配模型为基础,对边缘节点的计算资源、频谱资

源、缓存资源进行统一分配和优化。文献[24-25]中,作者分别以实现边缘计算卸载的时延最小化或者端设备功率损耗最小化为目标,进行相应约束条件下的单用户场景和多用户场景下的计算卸载。上述研究为我们求解边-云协同的传感制图框架奠定了理论基础。

## 7.2 边缘缓存

边缘缓存把数据缓存到网络边缘<sup>[26]</sup>,可以有效地提高网络频谱利用效益,缓解无线网络的资源紧张。边缘缓存可以提供时延更低、可靠性更高的数据服务。数据存储在更靠近用户;因此,用户请求可不必经过复杂的网络传输交由核心网处理,可直接由边缘缓存节点处理并回传给用户,将大大地减少回程数据链路传输所需的时间,分载回程网络流量。文献[27]中,作者把车联网下的车辆服务数据缓存建模为整数线性规划问题,以优化缓存内容的预先放置,但是它的模型是基于未来路径已知的假设,这对于实际的无人驾驶场景通常难以实现。文献[28]针对5G网络应用,利用收集到的用户情境信息(如浏览历史、位置信息),通过边缘智能(如机器学习)来预测服务数据在未来时刻的时空分布,预先将流行度高的内容缓存到边缘节点,实现用户数据服务体验的提升。出于相同的目的,文献[29]中,作者以一阶马尔可夫模型对车辆移动轨迹进行预测,并使用熵来量化预测结果的不确定性,以此为基础确定预先缓存方案。文献[30]的作者提出了一种多

层次的协作缓存方法,将缓存内容边缘放置问题建模为最优化问题,以设备容量、链路容量、缓存容量等为参数求解缓存内容在边缘节点的分布。在缓存数据分发方向,文献[31]的作者提出首先将热门请求数据随机缓存在移动设备上,然后使用索引编码对数据进行广播分发,可同时满足多用户的内容需求。文献[32-33]的作者提出利用移动设备间的设备到设备(D2D)直通,进行内容的分发与共享。文献[34]对用户的信息需求模式进行了分析,认为5G网络数据请求具有一定的可预测性,利用这种可预测性在感知应用情境的同时,制定面向社交网络的边缘缓存。上述边缘缓存的研究并考虑无人驾驶的应用场景,直接应用于高精地图数据缓存时效果可能难以保证。

## 7.3 实时交通数据感知

实时准确的交通数据是改善城市交通状况的信息基础,无人驾驶场景中以信息进行决策,实时交通数据感知显得更加重要。文献[35]中,作者提出利用车辆的高速移动性和多种传感信息进行全局环境感知,把采集到的数据直接汇聚到云中心进而提供智慧交通、智慧城市等数据服务。但是此框架下均限定车辆以特定模式进行数据采集,且传感数据类型简单。文献[36]中,作者构建了一个可以覆盖全路网的交通信息动态更新模型,自动地完成交通信息数据的收集、聚合、处理和传输。文献[24-25]中,作者利用外部数据、车间协作等信息评估交

通状况、感知交通拥塞程度。进一步地,文献[37]中,作者提出在车辆通过车间协作完成交通状况感知之后,还利用线性最小二乘法对下一阶段进行预测,求解短期交通状况的变化情况。文献[20]中,作者提出在发生交通拥堵时,对异常交通流量进行分布式聚类,实现感知数据的聚合。上述研究试图解决交通信息的多节点协作感知,但并未讨论感知信息的发布施效,并未闭环感知-服务-反馈环节。文献[38]中,作者提出一种预警传播机制,根据事故的严重程度将警告信息约束在相应的兴趣区域范围内。文献[39]提出了一个车载信息传输协议,通过此协议车辆向下游车辆查询前方实时交通状况。面向未来无人驾驶环境中,文献[23]中,作者提出了名为CarSpeak的协作感知与通信系统为无人驾驶提供实时数据支持。它利用八叉树结构来存储3D点云数据可实现车间感知数据共享时自适应的多分辨率选择。此外,它还还为车间感知数据共享提供了一种数据接入控制协议,通过竞争实现信道资源分配。文献[40]中,作者提出利用车辆到车辆(V2V)通信来获知交通事件信息,并结合自身评估信息的时空相关性,做出相应的驾驶行为调整以缓解交通拥塞。

## 8 结束语

边缘计算将服务下沉到了网络边缘,这为高精地图提供了业务本地化和业务近距离部署的条件。边缘地图服务更加靠近车辆和数据,

可以很大程度上减少网络交互和服务交付的时延,有效地满足无人驾驶超低时延的需求。相比云端集中式的地图数据处理,边缘地图节点配置的业务计算支撑和地图数据存储,可在时空有效范围内对传感数据进行处理、分析、缓存、发布,并推送有价值的处理结果至云地图中心完成后续决策。这可以极大地节省了链路数据资源,提高地图服务业务效率,支持更优的用户体验。因此,边缘高精地图服务是无人驾驶今后实现普及的基础支撑。值得注意的是:高精地图服务应该具备全路况的地图提供能力,不仅要覆盖高速公路这种相对简单的路况,还更应该覆盖复杂路况区域,比如旧城区、多口路口等等。这些区域干扰因素更多,需要更加准确地理解周边环境,这部分的边缘制图能力才是高精地图真正的挑战所在。同时,随着无人驾驶的逐步普及,为应对各类突发状况,高精地图需要更多的半动态数据以及动态数据,这大大提高了在边缘地图服务中对数据实时性的要求。不同层次领域的地图数据时效不同,例如道路的几何形状很少会发生变化,不需要进行实时更新,而交通信息则是实时更新。因此,未来需要进一步地定义更高效的(半)自动化的边缘地图制作流程,以完成对地图的更新和对无人驾驶的辅助。

#### 参考文献

- [1] 刘少山,唐洁.第一本无人驾驶技术书[M].北京:电子工业出版社,2017  
 [2] NEDEVSCHI S, POPESCU V, DANESCU R, et al. Accurate Ego-Vehicle Global Localization at

- Intersections through Alignment of Visual Data with Digital Map[J]. IEEE Transactions on Intelligent Transportation Systems, 2013, 14(2): 673-687. DOI:10.1109/tits.2012.2228191  
 [3] 贺勇.基于高精地图的GPS导航方法研究[D].上海:上海交通大学,2015  
 [4] SUGANUMA N, UOZUMI T. Precise Position Estimation of Autonomous Vehicle Based on Map-Matching[C]//2011 IEEE Intelligent Vehicles Symposium (IV). Germany: IEEE, 2011: 296-301. DOI:10.1109/IVS.2011.5940510  
 [5] LI H, NASHASHIBI F, TOULMINET G. Localization for Intelligent Vehicle by Fusing Mono-Camera, Low-Cost GPS and Map Data [C]//13th International IEEE Conference on Intelligent Transportation Systems. Portugal: IEEE, 2010: 1657-1662. DOI:10.1109/ITSC.2010.5625240  
 [6] RESS C, ETEMAD A, KUCK D, et al. Electronic Horizon—Providing Digital Map Data for ADAS applications [J]. Madeira, 2008, (3):40-49  
 [7] LEVINSON J, MONTEMERLO M, THRUN S. Map-Based Precision Vehicle Localization in Urban Environments [M]. Cambridge: MIT Press, 2007  
 [8] SCHREIBER M, KNÖPPEL C, FRANKE U. LaneLoc: Lane Marking Based Localization Using Highly Accurate Maps[C]//2013 IEEE Intelligent Vehicles Symposium (IV). Australia: IEEE, 2013: 449-454. DOI:10.1109/IVS.2013.6629509  
 [9] GUO H Z, MEGURO J I, KOJIMA Y, et al. Automatic Lane-Level Map Generation for Autonomous Robotic Cars and Advanced Driver Assistance Systems Using Low-Cost Sensors[C]//IEEE International Conference on Robotics & Automation ICRA. China: IEEE, 2014  
 [10] SUTARWALA B Z. GIS for Mapping of Lane-Level Data and Re-Creation in Real Time for Navigation [D]. Riverside: University of California (Master's thesis), 2010  
 [11] BENDER P, ZIEGLER J, STILLER C. Lanelets: Efficient Map Representation for Autonomous Driving[C]//2014 IEEE Intelligent Vehicles Symposium Proceedings. USA:IEEE, 2014: 420-425. DOI:10.1109/IVS.2014.6856487  
 [12] JIAO J L. Machine Learning Assisted High-Definition Map Creation[C]//2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC). Japan: IEEE, 2018: 367-373. DOI:10.1109/COMPSAC.2018.00058  
 [13] SHI W S, CAO J, ZHANG Q, et al. Edge Computing: Vision and Challenges[J]. IEEE Internet of Things Journal, 2016, 3(5): 637-646. DOI:10.1109/ijot.2016.2579198  
 [14] 施巍松,孙辉,曹杰,等.边缘计算:万物互联时代新型计算模型[J].计算机研究与发展,2017,54(5): 907-924. DOI:10.7544/issn1000-1239.2017.20160941  
 [15] AHMED A, AHMED E. A Survey on Mobile Edge Computing[C]//2016 10th International Conference on Intelligent Systems and Control (ISCO). India: IEEE, 2016: 1-8  
 [16] European Telecommunications Standards Institute (ETSI). Mobile-Edge Computing Introductory Technical White Paper [EB/OL]. (2018-09-03) [2019-00-00]. https://portal.etsi.org/portals/0/tbpages/mec/docs/mobile-edge\_computing\_-\_introductory\_technical\_

- white\_paper\_v1%2018-09-14.pdf  
 [17] 3GPP. Procedures for the 5G System (Release 15): 3GPP TR 23. 502[S]  
 [18] MENDEZ D, LABRADOR M A. Density Maps: Determining Where to Sample in Participatory Sensing Systems[C]//2012 Third FTRA International Conference on Mobile, Ubiquitous, and Intelligent Computing. Canada, 2012: 35-40. DOI:10.1109/MUSIC.2012.14  
 [19] SONG Z, NGAI E, MA J, et al. Incentive Mechanism for Participatory Sensing under Budget Constraints[C]//2014 IEEE Wireless Communications and Networking Conference (WCNC). Turkey: IEEE, 2014: 3361-3366. DOI:10.1109/WCNC.2014.6953116  
 [20] DORNBUSH S, JOSHI A. Street Smart Traffic: Discovering and Disseminating Automobile Congestion Using VANET's[C]//2007 IEEE 65th Vehicular Technology Conference - VTC2007-Spring. Ireland: IEEE, 2007: 11-15. DOI:10.1109/VETECS.2007.15  
 [21] ZHANG Q, ZHAO J H. A Model for Automatic Collection and Dynamic Transmission of Traffic Information Based on VANET[C]//2012 15th International IEEE Conference on Intelligent Transportation Systems. USA: IEEE, 2012: 373-378. DOI:10.1109/ITSC.2012.6338711  
 [22] TERROSO-SAENZ F, VALDES-VELA M, SOTOMAYOR-MARTINEZ C, et al. A Cooperative Approach to Traffic Congestion Detection with Complex Event Processing and VANET[J]. IEEE Transactions on Intelligent Transportation Systems, 2012, 13(2): 914-929. DOI:10.1109/tits.2012.2186127  
 [23] KUMAR S, SHI L, AHMED N, et al. Carspeak: A Content-Centric Network for Autonomous Driving[C]//Proceedings of the ACM SIGCOMM 2012 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication. USA: ACM, 2012: 259-270  
 [24] BAUZA R, GOZALVEZ J. Traffic Congestion Detection in Large-Scale Scenarios Using Vehicle-To-Vehicle Communications [J]. Journal of Network and Computer Applications, 2013, 36(5): 1295-1307. DOI: 10.1016/j.jnca.2012.02.007  
 [25] GRAMAGLIA M, CALDERON M, BERNARDOS C J. ABEONA Monitored Traffic: VANET-Assisted Cooperative Traffic Congestion Forecasting[J]. IEEE Vehicular Technology Magazine, 2014, 9(2): 50-57. DOI:10.1109/mvt.2014.2312238  
 [26] LIU D, CHEN B Q, YANG C Y, et al. Caching at the Wireless Edge: Design Aspects, Challenges, and Future Directions[J]. IEEE Communications Magazine, 2016, 54(9): 22-28. DOI:10.1109/mcom.2016.7565183  
 [27] MAURI G, GERLA M, BRUNO F, et al. Optimal Content Prefetching in NDN Vehicle-To-Infrastructure Scenario [J]. IEEE Transactions on Vehicular Technology, 2017, 66(3): 2513-2525. DOI:10.1109/tvt.2016.2580586  
 [28] ZEYDAN E, BASTUG E, BENNIS M, et al. Big Data Caching for Networking: Moving from Cloud to Edge [J]. IEEE Communications Magazine, 2016, 54(9): 36-42. DOI:10.1109/

➔ 下转第 81 页

# 边缘计算中数据驱动的智能应用:前景与挑战

## Data-Driven Intelligent Application in Edge Computing: Prospects and Challenges



秦永彬/QIN Yongbin<sup>1</sup>, 韩蒙/HAN Meng<sup>2</sup>, 杨清亮/YANG Qingliang<sup>2</sup>

(1. 贵州大学贵州省公共大数据重点实验室, 贵州 贵阳 550025;

2. 肯尼索州立大学数据驱动智能研究实验室, 美国 乔治亚州 30066)

(1. Guizhou Provincial Key Laboratory of Public Big Data, GuiZhou University, Guiyang 550025, China;

2. Data-driven Intelligence Research (DIR) Laboratory, Kennesaw State University, GA 30066, USA)

**摘要:** 针对边缘计算中数据驱动的智能应用这一子领域进行了全面的综述和比较。创新性地提出面向智能应用的边缘体系架构, 并对目前已经存在的智能应用的最新成果进行总结。通过对相关研究工作的总结和分析, 提出将人工智能与边缘计算相结合的方法来面对全新的机会和挑战, 并对如何衡量、面对并有效利用这些机会与挑战提出了进一步的思考和展望。

**关键词:** 边缘计算; 数据驱动; 智能应用; 人工智能

**Abstract:** In this paper, a comprehensive overview and comparison of the sub-areas of data-driven intelligent applications in edge computing is given. An edge architecture for intelligent applications is proposed and the latest achievements of existing intelligent applications are summarized. It points out that integrating artificial intelligence and edge computing will bring new opportunities and challenges. Further thoughts and prospects are then proposed on how to measure, face and effectively use these challenges and opportunities.

**Key words:** edge computing; data driven; intelligent applications; artificial intelligence

DOI: 10.12142/ZTETJ.201903010

网络出版地址: <http://kns.cnki.net/kcms/detail/34.1228.TN.20190611.1437.002.html>

网络出版日期: 2019-06-11

收稿日期: 2018-12-23

随着物联网、5G网络的全方位覆盖, 大数据、人工智能等技术也逐渐渗透到了各行各业以及人们的日常生活中, 智慧城市、移动支付等诸多新型服务模式和业务不断出现, 帮助着经济社会转型升级的同时, 也推动了人类生活的发展。近几年的物联网设备连接数也呈现出线性增长趋势, 据互联网业务解决方案集团预测: 2020年的物联网设备数量将达到500亿, 具体如图1

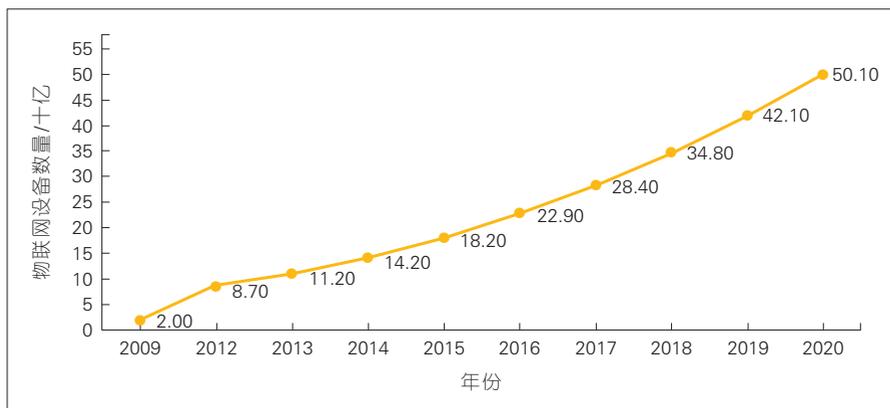
所示。

在此万物互联的大数据时代背景下, 一方面, 每天产生的数据量剧增; 另一方面, 物联网等应用背景下的数据在地理上分散, 对响应时间和安全性提出了更高的要求。云计算, 作为当今主要的计算技术之一, 凭借其便捷的、按时按需的计算模式, 在过去的10年里, 成功地缓解了日益增长的数据所带来的存储、管理等问题。传统的云计算虽然为大数据处理提供了高效的计算平台, 但是目前网络带宽的增长速度

远远赶不上数据的增长速度, 网络带宽成本的下降速度要比中央处理器(CPU)、内存这些硬件资源成本的下降速度慢很多, 同时复杂的网络环境让网络延迟很难有突破性提升。所以, 传统的云计算不能满足对响应时间和安全性的高要求。单纯依靠云计算这种集中式的计算处理方式, 不足以支持以物联网感知为背景的应用程序运行和海量数据处理。

在这种应用背景下, 边缘计算应运而生, 它与现有的云计算集中

基金项目: 贵州省公共大数据重点实验室开放课题基金(No.2018BDKFJJ002)



▲图1 物联网设备增长趋势

式处理模型相结合,能有效解决云中心和网络边缘的大数据处理问题。边缘计算是指数据或任务能够在靠近数据源头的网络边缘侧进行计算和执行计算的一种新型服务模型。这里所提的网络边缘侧是从数据源到云计算中心之间的任意功能实体。

相比于云计算,边缘计算具有以下明显的优点:

- 边缘计算可以在网络边缘处理大量临时数据从而减轻网络带宽与数据中心的压力。

- 边缘计算在用户附近提供服务,近距离服务保证了较低的网络延迟,简单的路由也减少了网络的抖动,使边缘服务比云服务有更强的响应能力。

- 边缘计算则为关键性隐私数据的存储与使用提供了基础设施,提升数据的安全性。

边缘计算因为其突出的优点,满足未来万物互联的需求,从2016年开始迅速升温,引起全球的密切关注。目前,关于边缘计算的研究有很多,如关于边缘计算数据安全与隐私保护的研究<sup>[1]</sup>、关于移动边

缘计算通信方式的研究<sup>[2]</sup>,也有关于移动边缘计算架构的研究<sup>[3]</sup>等,但是目前还没有关于边缘计算中数据驱动的智能应用的研究。

不同于目前关于边缘计算的研究综述,本文中我们第一次从边缘计算与数据驱动的智能应用相结合的角度做全面的研究综述,以期对这一极具前景的领域提供全面的概览以及展望。物联网终端产生的“海量级”数据所蕴含的巨大潜在价值将会促使由数据驱动的智能应用的飞速发展。而在边缘计算模式下,搭载在网络边缘侧由数据驱动的智能应用将会表现出更强的响应能力。因为数据可以在靠近源头侧进行处理,在用户附近提供服务,近距离服务保证了网络延迟,简单的路由减少了网络抖动。综合以上2点,有理由相信在边缘计算模式下,由数据驱动的智能应用将会得到更快、更大的发展,尤其会在提升个性化客户体验、优化客户互动、提高运营效率、启用新的业务模型方面发挥重要作用。数据驱动的智能将在深度和广度多个维度下全面影响未来的社会和世界。

本文中,我们针对于边缘计算中的数据驱动的智能应用进行详尽的归类、分析和总结,以期为边缘计算领域的研究人员,尤其是针对于边缘计算中数据驱动的智能应用有关的研究人员提供一个全面的参考。本文的贡献为:(1)率先提请学术与工业界对边缘计算中基于数据驱动的智能应用这一子领域进行关注;(2)广泛地分析、分类并总结了诸多具有未来发展潜力的边缘计算研究方向;(3)对具有潜力的多个未来研究方向提出了相应的技术展望。希望可以同各个相关领域的专家一起将边缘计算拓展出更多有理论价值和实用价值的研究成果。

## 1 面向智能应用的边缘体系架构

本文中,我们定义的是面向智能应用的边缘计算体系架构。边缘指的是网络中作为计算节点的物联网设备在地理上分布于企业或其他网络的边缘。边缘计算是指数据或任务能够在靠近数据源头的网络边缘侧执行计算的一种新型服务模型,其动机是将服务器资源、数据分析以及人工智能等计算更靠近数据产生和收集的源头。如图2所示,面向智能应用的边缘计算体系架构主要包括基础智能应用、边缘网络智能应用以及移动智能应用这3个层次。

### 1.1 基础智能应用

基础智能应用是指部署在核心基础设施上的智能应用。核心基础设施为网络边缘设备提供核心网络

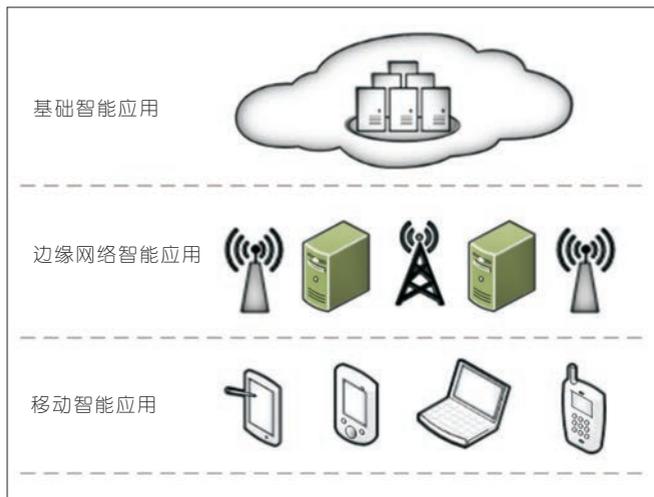


图2 面向智能应用的边缘计算体系架构

接入,核心网络主要包括互联网络、移动核心网络、数据中心等。数据中心为云计算服务提供硬件基础,通常会设立在人口稀少、能源利用率相对较低且不易发生灾害的区域,现代的数据中心通常由成千上万的服务器相互连接组成。数据中心的能源管理是基础智能应用的方向之一。配电单元使用智能来帮助企业或组织绘制能量图;网络化的配电单元,比如美国电力转换公司(APC)的 Switched Rack AP8000 系列和 CyberPower 的监测控制设备,可以提供实时电源监测控制和温度/湿度感应,它们也控制插座级别的电源,以便对设备(例如单个机架式服务器)进行粒度循环。数据中心基础设施管理(DCIM)指从数据中心设施到服务器和设备级别的粒度报告软件。DCIM 工具可帮助信息技术(IT)专业人员进行容量规划、系统库存控制和生命周期管理,工作负载平衡和服务器整合,具体包括:关闭空闲服务器与监测控制和提高系统恢复能力,以及其他基于洞察力的工作举措<sup>[4]</sup>。人工智能

可以发挥重大作用:提供深入的洞察力来支持数据中心专家,例如热优化引擎核心的人工智能启用多站点热分析工作正在进行中。这一突破为用户提供了强大的仿真工具,可以提供有关现场无效率或风险区域的实时建议。

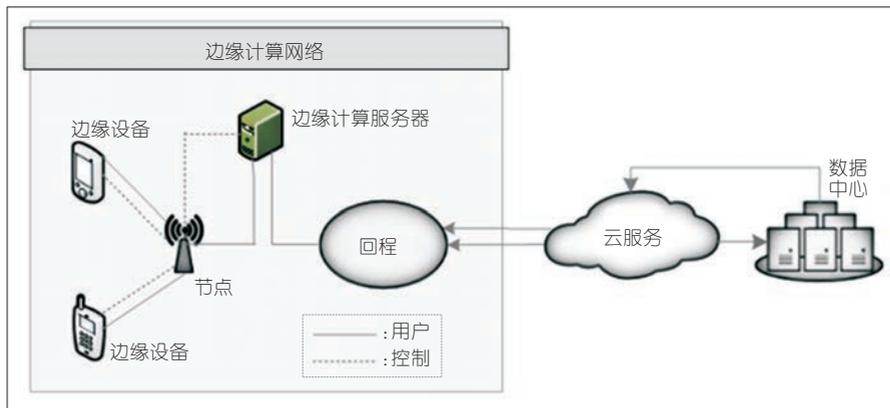
### 1.2 边缘网络智能应用

边缘网络智能应用指部署在边缘网络节点的智能应用平台。边缘网络位于集中式网络外围,为中央或核心网络提供信息,包括无线网络、移动中心网络和互联网络。边缘网络节点包括蜂窝网络基站、无线网络接入点、路由器等。

目前,许多应用程序依赖于远程数据中心托管的数据和服务,这会加大网络负载,因为数据需要在连接移动设备和远程数据中心的互联网内上传和下载。激增的数据流量会导致网络拥塞和延迟。边缘网络旨在通过将计算任务从互联网转移到边缘设备来减少网络压力和延迟,所以边缘设备需要提供计算和存储服务而不是仅充当移动设备的

接入点。

图3 描述了边缘网络的拓扑结构。边缘设备连接到节点转换无线电信号,使得信号可以发送到有线网络和核心网络。边缘服务器通过物理连接部署在节点附近,并且通过在服务器内部循环流量来进一步处理数据。通过以上方式,边缘服务器能够直接在网络边缘处理特定用户需求,而不需要将所有流量转发至远程 Internet 服务。边缘服务器可以直接处理请求并直接响应边缘设备,或者转发请求到远程数据中心请求云服务。到目前为止,移动边缘服务器尚未有部署在蜂窝网络中的实际应用,主要的研究仍仅从理论角度讨论移动边缘服务器的概念。与此概念相关,M. SATYANAYANAN 等人讨论了 Cloudlet 概念。Cloudlet 是一个可信且资源丰富的主机或机群,它部署在网络边缘,与互联网连接,可以被周围的移动设备所访问,为设备提供服务。Cloudlet 可以将资源密集型任务卸载到其他资源上,以提高任务执行速度和延长电池寿命。Cloudlet 软件栈共分3层:第1层是操作系统和缓存,第2层是虚拟化层,第3层是虚拟机实例,负责运行移动设备卸载的应用,在虚拟机中运行应用可以弥补移动设备和 Cloudlet 平台的差异性。Cloudlet 部署在网络边缘,可以被周围的移动设备访问。Cloudlet 平台降低了网络延迟,加快了许多复杂应用的响应速度。移动数据流量持续增长、智能手机的普遍使用和物联网的实现都对移动网络基础设施提出了更



▲ 图3 边缘网络拓扑结构

高的要求,但是移动网络基础设施在最初设计时并未考虑这种新形势下的要求。通过在移动网络边缘加入更多的智能,可以使移动网络运营商能够优化其基础设施,从而应对前所未有的数据流量,加速数据服务,提升用户体验。在网络边缘部署智能可增强延迟关键应用程序,优化运营商网络流量,提供可分析大量信息的分布式计算环境,且可以比传统数据中心环境更快地完成此任务,具体如图4所示。

### 1.3 移动智能应用

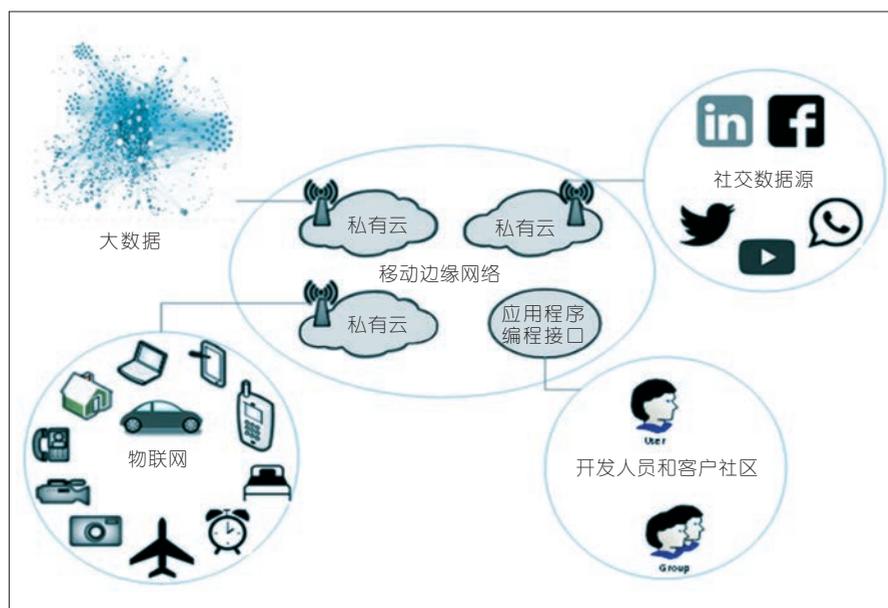
移动智能应用是指部署在移动终端上的智能应用。移动终端包括连接到边缘网络中的所有类型的设备,具体包括智能移动电话、平板电脑、物联网设备等。随着万物互联趋势的加剧,移动终端设备也呈现快速增长的趋势。

边缘计算在蜂窝网络中最明显的优势是减少端到端的延迟。当数据包不必在通过核心网络传到互联网上的应用程序服务器时,应用程序可以提供具有强大、稳定和已知延迟边界的实时服务。边缘计算使

得任务卸载在很多情况下是可行的,因为目前的无线带宽比可用的互联网带宽高。将计算密集型任务卸载到边缘服务器执行这一方式使得越来越多新兴的移动应用程序从中受益。例如,增强现实(AR)应用程序,它将现实世界的应用场景与虚拟信息高度集成,生成能被人类感官感知的信息,来达到超越现实的感官体验。AR可以在移动设备(智能手机、平板电脑、智能眼镜等)上使用,支持新的应用和服务,比如

3D观影、虚拟游戏等。如图5所示,AR应用有5个关键组件<sup>[5]</sup>:视频源(从移动摄像机获取原始视频帧)、跟踪器(跟踪用户的位置)、映射器(用于构建环境模型)、对象识别器(用于标识环境中的已知对象)和渲染器(用于准备已处理的帧以进行显示)。在这些组件中,视频源和渲染器应该在本地执行,而其他计算密集型组件,如跟踪器、映射器和对象识别器可以卸载到边缘服务器或云执行。这样,终端设备用户就会享受诸如延迟减少和设备节能等带来的优势。

除AR应用程序外,还有人脸识别应用程序很适用于边缘计算模式。如图6所示,人脸识别应用程序通常由5个主要的计算机组件组成:图像采集、人脸检测、预处理、特征提取和分类<sup>[6]</sup>。在这些组件中,图像获取组件需要在移动设备上执行以支持用户界面,但是可以把其他包含复杂计算的组件卸载到边缘



▲ 图4 移动边缘网络

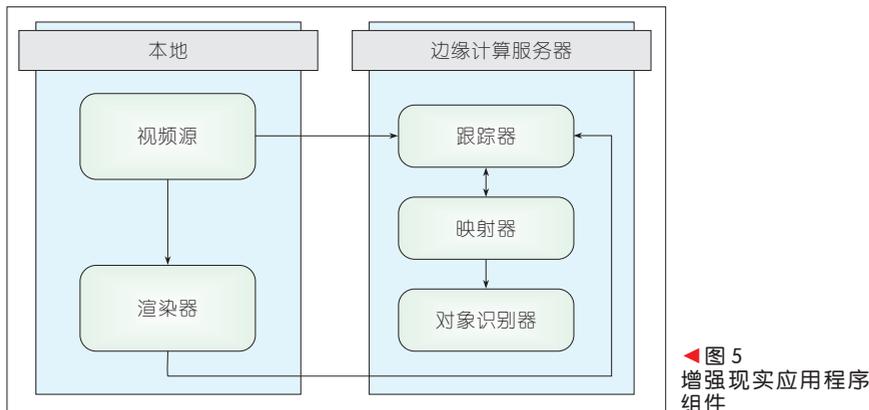


图5 增强现实应用程序组件

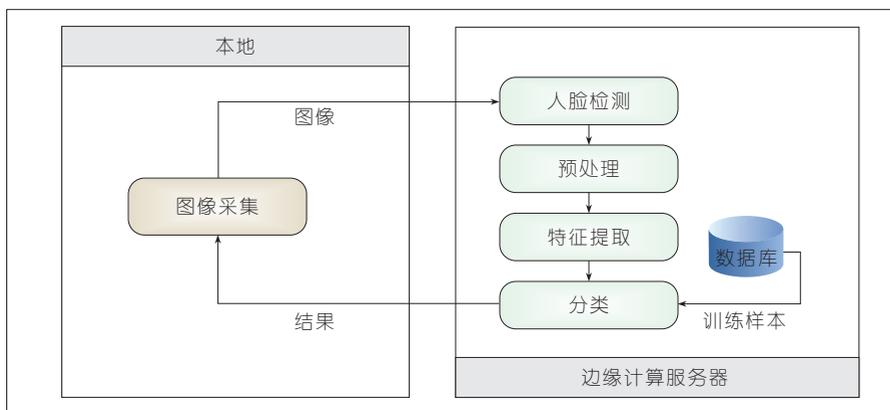


图6 人脸识别应用程序组件

服务器处理,比如信号处理和机器学习算法。

总体来讲,边缘计算范式可以给移动运营商、服务提供商和终端移动用户带来很多好处。如果从最终用户的角度去看,面向消费者的服务因为对最终用户直接有利,所以最受关注。通常,用户主要通过计算卸载获益,这使得用户能够在移动终端运行新兴的应用。

## 2 现有智能应用

### 2.1 无人驾驶汽车

无人驾驶汽车关键技术包括:环境感知、导航定位、路径规划、决策控制。

环境感知是无人驾驶汽车的关键技术之一。无人驾驶汽车通过环境感知模块辨别车内信息和周围环境信息,为其行为决策提供信息支持。对于车内信息,比如车辆自身的速度、加速度、位置等,主要采用驱动电机、电子罗盘、陀螺仪等传感器进行感知。对于周围环境信息如街道范围、信号灯、路标等信息的识别,主要采用雷达等主动型测距传感器为主、被动型测距传感器为辅的方法实现。对周围环境感知的难点在于对模糊标志的识别,随着计算机视觉和机器学习等技术普及和深入,感知技术近几年有了很大的提高。传感器带回来的信号带有大量的噪音和不确定性,需要采用

传感器融合和定位技术过滤无用信息,将客观世界进行3D或4D还原。可见无人驾驶汽车环境感知技术是以传感器等硬件为基础,大量收集信息,融合处理数据,最终为行为决策提供依据。

### 2.2 智能数据库

智能数据库包含人工智能组件,可以为搜索的智能操作提供帮助,具有表示知识的方式,并且基于连接主义神经网络模型。连接主义神经网络模型指将新信息与已知的类似信息相关联的模型<sup>[7]</sup>。智能数据库与用户交互以确保返回的项目中包含尽可能相关的一些信息。

传统的数据库全文检索技术是使用字符串搜索来匹配用户输入的关键词和短语,但是关键词和短语的匹配并不能保证该文件的内容在给定的上下文中是最相关的。相对于传统数据库,智能数据库提高了查询的灵活性。例如,智能数据库可以根据用户键入的问题,提供根据可能性(从最高到最低)排列的命中列表,返回的结果数据包含对该问题的有用答案。人工智能可以纠正用户的一些输入错误(比如拼写错误),也可以提供关键字的同义词和反义词等供用户进行选择。

智能数据库搜索可以限于特定领域,比如商业新闻、图像、信息技术等,也可以限于某些语言、地理区域或历史时期。智能数据库还可以为用户提供最近的搜索记录;因此如果用户需要优化搜索或重新搜

错,则不必重新开始。

### 2.3 金融领域智能化

人工智能的最大进步尚未到来,但大数据与机器学习算法的结合已经为金融世界的日常运营带来好处。人工智能正在逐渐改变金融业务核心,其影响主要体现在信用评级(信誉评级)、市场研究、个人助理、资产管理和欺诈检测等方面。大多数银行已经加入人工智能的行列,再过几年认知系统和人工智能的广泛采用有望提高全球收入。国际数据公司(IDC)的一份报告预测:到2020年收入增长可能达到470亿美元。

总部位于美国的富国银行2017年初开始试行人工智能驱动的Facebook聊天机器人。这个虚拟助手能够与用户沟通,响应来自客户的查询,例如查询他们帐户中的当前余额,甚至找到最近的银行自动取款机(ATM)。随着名称为Erica的智能虚拟助手的亮相,美国银行公司大胆推进了人工智能技术。Erica是一个利用“预测分析和认知信息”的聊天机器人,被称为全球最大的支付和金融服务创新活动,为该公司超过4500万客户提供财务指导。作为移动银行体验的一个组成部分,Erica旨在为客户提供“全天候访问并执行日常交易”,同时预测每个客户的独特财务需求并通过提供明智的建议帮助他们实现财务目标。花旗银行对Feedzai进行了战略性投资。Feedzai是一家领先的全球数据科学企业,可以实时识别和消除所有商业渠道中的欺诈行

为,包括在线和面对面的银行业务。通过对大量数据的持续快速评估,Feedzai可以进行大规模分析,识别出欺诈或可疑活动,并迅速提醒客户。该服务还协助支付提供商和零售商监测控制和保护与其公司相关的金融活动。为防止商业中出现的对客户的潜在威胁,Feedzai利用“基于机器的学习”来评估“大数据”和其潜在的欺诈活动。

### 2.4 智慧城市

Forrester将智慧城市定义为:利用智能计算技术使城市的关键基础设施组件和服务更加智能、互联和高效。关键基础设施组件和服务包括:城市管理、教育、医疗保健、公共安全、房地产、交通和公用事业。

关于智能计算,Forrester将其定义为<sup>[8]</sup>:集成硬件、软件和网络的新一代技术,为IT系统提供对现实世界的实时感知和高级分析,帮助人们针对优化业务流程和业务资产负债表结果的备选方案和行动做出更加明智的决策。智慧城市旨在帮助解决城市化进程中遇到的各种问题,例如,资源稀缺、基础设施不足和恶化、能源短缺和价格不稳定、全球环境问题和人类健康问题等。智慧城市需要实时感知、分析整个城市的各项相关信息,会产生大量的原始数据,且这些数据大部分存储在本地,不利于查找和分析,因此,需要一种有效的解决方案进行数据处理。TANG B.<sup>[9]</sup>等人提出一种大数据分析框架,该框架使用了边缘计算技术,共分为4层:第1层为传感器网络,由分散在城市的各个传

感器组成,负责生成原始数据;第2层为边缘节点,每个边缘节点都控制一组本地传感器,可以根据预先设定好的模式分析处理原始数据;第3层为中间计算节点,每个中间计算节点控制一组边缘节点,依据边缘节点上传的信息,并且综合时空信息来判断潜在突发事件,且能控制下层设备做出应急反应;第4层为云计算中心,对全市范围进行长期监测控制和分析。该框架因采用边缘计算技术,显著减少了上传到云端的数据量,减少网络延迟,保证其高效运行。

## 3 边缘计算中数据驱动智能应用的发展与方向

### 3.1 智能助手

智能助手是一款应用程序,主要通过内建的自然语言或其他易用的交互用户界面来辅助人们处理基本任务。智能语音助手可以基于用户输入的数据帮助回答用户询问,实际上可以帮助人们实时解决问题从而增强人类生产力的机器人。最初,只有智能手机和智能扬声器采用语音助手,现在智能手表、汽车、笔记本电脑、家用电器等越来越多的设备采用语音助手,它正在成为人们与技术设备日常交互的普遍方式。关于启用语音助手的问题,如果设备不停地录制音频并将其发送到云端以验证是否有人向他们发送指令,这不仅涉及隐私问题,也是在浪费能源、计算和网络资源。将所有的数据发送到云端并在服务器端处理将会带来严重的延迟,也会降

低整个系统的可用性。今天的语音接口通常使用关键字或“唤醒字”检测,专用一小部分边缘计算资源(即在设备本身或“边缘”完成的计算)来处理麦克风信号,同时系统的其余部分保持空闲。这是一种节能方法,对于帮助延长便携式电池供电设备(包括智能手机和可穿戴设备)的使用时间尤为重要。检测到关键字以后,设备开始主动监测听查。此时,系统是否能准确解释语音命令很大程度上取决于语音的“干净”程度,也就是说设备在嘈杂的环境中很难准确解释语音,比如会议室、街道、电影院等。

许多边缘计算技术有助于将用户的声音与其他周围的声音分开。例如,波束形成技术处理来自设备中的多个麦克风的音频,以将听力集中在用户正在讲话的方向上,如虚拟定向麦克风。如果用户四处移动,则设备上运行的语音跟踪算法可以调整来自麦克风的信号之间的平衡,因此焦点跟随语音源。高级语音设备还处理来自麦克风阵列的输入,以抑制用户通话中的环境噪声,类似于在降噪耳机中操作的方式。智能扬声器还使用设备上的回声消除技术来实现“插入”功能,这可以抑制麦克风信号中的音乐和其他扬声器声音,从而帮助智能扬声器即使在大声播放音乐时也能接收语音命令。在支持语音的小工具中增加边缘计算功能还支持使用设备上人工智能的创新功能。例如,离线命令允许在互联网连接不可用时进行设备语言处理和基本语音指令的执行。此功能已在智能手

机中广泛使用,即使设备处于飞行模式或超出覆盖范围,也可帮助用户设置警报和提醒。具有高级边缘计算能力的设备还可以执行语音生物识别以进行用户认证。语音设备对优异边缘处理能力的需求推动了异构计算架构的采用:语音设备将各种引擎集成到单个片上系统中,将工作负载分配给最高效的计算引擎,从而改善性能、功效和成本效益,支持包含语音接口的各种设备。

### 3.2 行业应用

人工智能在边缘计算有关方面的研究与工业潜力巨大。Tractica的一份报告估计:到2025年,人工智能边缘设备出货量将从2018年的1.614亿台增加到全球每年26亿台。就单位数量而言,顶级的人工智能边缘设备将包括手机、智能扬声器、个人电脑/平板电脑、头戴式显示器、汽车传感器、无人机、消费者和企业机器人以及安全摄像头。可穿戴式健康传感器、建筑物或设施传感器还将有更多人工智能。例如,关于智慧城市的共识是智慧城市利用物联网传感器、执行器和技术来连接整个城市的各个部分,要实现智慧城市的愿景就要充分融合物联网技术、人工智能和边缘计算。

这一领域应用前景广阔,如可以帮助解决交通拥堵问题、城市卫生等。关于交通拥堵问题,新加坡通过联网运输解决方案应对了巨大的交通挑战。其智能交通系统(ITS)已经形成了电子道路收费的先锋系统,随着交通量激增公路收费也随之增加。ITS还允许通过具

有全球定位系统(GPS)功能的出租车提供实时交通信息,并整合公共交通结构,同时也能够使公交车更加准时<sup>[10]</sup>。城市卫生条件也可以通过智能技术改善,如使用互联网连接的垃圾桶或物联网支持的车队管理系统进行废物收集和垃圾处理,或者是将传感器技术应用于智能废物容器,(它会自动感知什么时候垃圾到达容器的顶部,从而及时通知城市环卫部门收集垃圾)。在医疗设备领域特别是医疗设备急需边缘情报,例如,手术室中的关键数据必须及时处理,以便为医生或医生提供正确的信息。另外,受益的潜在领域是机器视觉,主要使用摄像头和可视化分析。例如,可以将摄像机放置在配送设施中,以便在瞬间监测和管理在卡车和托盘之间移动的货物。

另一个受益的领域是农业。天气状况对于农业来说至关重要,但它又是无法控制的。一家总部位于澳大利亚的农业技术公司The Yield,它使用传感器、数据和人工智能帮助农民做出有关天气、土壤和植物条件的明智决策。

其他的应用还包括新闻。像AP和雅虎这样的公司正在使用人工智能来准备与体育和选举相关的简单报告,而这些报告如果手动完成则需要花费大量时间。据引入Wordsmith的Automated Insights称:包括房地产和电子商务在内的许多数据驱动实体都在使用这个平台。在线零售商店使用人工智能的方式是根据客户过去的购买记录或者放在搜索框中的项目向客户推荐购买

的内容;另一种方法是提供聊天机器人,以寻求指导或解决查询。另外还有智能家居,如恒温器和智能灯,可以在一天中的不同时间以用户想要的方式设置家的温度,还可以利用用户在不同时间段优选的暗淡和亮度来修改室内灯光效果。

## 4 机遇与挑战

边缘计算中数据驱动的智能应用将有广泛的前景,同时也面对着诸多的技术和实现环境挑战,本节将从以下5个方面对各个机遇挑战进行分析。

### 4.1 边缘计算中的数据分析

无论是在物联网的终端设备,还是手机、平板等移动端的诸多应用中,智能应用的基础均是对用户数据的收集、分析和建模,从而为客户提供个性化的智能推荐。

想要提取出有价值的信息,首先分析的数据量要大,所需的收集时间要充足;其次,数据分析结果要在时效期内,但也不可以被过早地分析,因为过早分析可能会导致丢失一些重要信息。如何平衡数据分析的价值和时效性是一项挑战。另外,用传统的机器学习的方法去学习用户行为,需要有足够的历史数据,但是在边缘计算的环境下,更多的是实时数据,怎样利用实时数据并且进行快速的分析和预测,将是另一个挑战。

如果可以有效地达成边缘端数据与全局数据的融合从而进行深入分析,必将更高效地推动更多的数据流应用,以及基于本地与全局相

互通的个性化智能推荐与应用。

### 4.2 边缘计算中数据智能应用的隐私安全

计算机安全、网络安全以及最新出现的越来越紧迫的数据安全与隐私已经引起广泛关注。美国的Facebook以及多家利用用户数据驱动应用的服务商或设备商已因安全及隐私问题在2018年遭受重创。安全与隐私问题也必将受到越来越广泛的关注。而在网络边缘设备产生的大量数据则会涉及更多的包括地理位置、行动轨迹等隐私信息,安全问题更为突出。一方面如何将传统的隐私保护方案和边缘计算环境下的数据处理特性相结合以实现多样化的用户隐私保护需求将成为一个新的挑战;另一方面,如何建立新的隐私保护模型和方法,适应于边缘设备的特性将成为另一个挑战。

边缘网络具有数据边缘性的特点,由此带来的便是数据不会完全被回传至中心网络及中心数据库中。事实上,实现终端数据在边缘设备端完成有效的匿名化和保护,将驱动一系列新的隐私保护策略,例如在数据产生时就进行有效地处理和保护,就不必再担心因中心数据非法接入所带来的各种风险和安全隐患了。

### 4.3 边缘计算中数据智能应用的接入与移除

边缘服务器以分散的方式部署。移动用户的位置不固定并经常实时发生变化,因此移动终端设备和服务器之间的连接将受到很多限

制,这使得边缘计算中诸多智能应用的接入和移除必须要有创新的形式和特点。首先在接入这些应用时,以何种方式部署就是一个新的挑战,采用分布式的计算架构还是点对点的分布式结构都可能是有效的解决方案,但尤其适配于智能应用的有效方法仍待进一步探索。相应地,当需要移除边缘智能应用程序时则需要具备移动感知能力,以进行资源发现和资源切换。另外,还需要评估卸载决策因素,决定何时将数据卸载到移动边缘计算服务器。由于受到链路质量、噪音干扰和通信拥塞等诸多因素的影响,如何完成这一系列的复杂操作也将极具挑战。

在云计算的不断发展过程中,关于无服务器(serverless)类应用的发展和探索一直没有停息,这类探索也可能成为未来边缘计算中智能应用接入和移除的一个新机遇。针对如何在边缘端建构出轻量级的接入和移除方式,并充分利用无服务器的敏捷特点,很有可能开发出一种新的应用使用模型。这可能是即插即用、即走即弃的方式,也可能将个性化的数据以边缘设备做为轻量载体的方式,从而可以非常容易地在不同的环境场景下切换。

### 4.4 边缘智能应用准确性和有效性

对于很多行业,例如自动驾驶,错误的决定将会导致严重的后果,所以如何保证决策的准确性和有效性将会一直是一个极具挑战的研究方向。人工智能往往使用大量的时

间并利用大规模的历史数据去“学习”。如何利用现有技术例如迁移学习去节省学习时间或利用对抗学习去模拟生成数据进行决策,都将是极具研究价值的可能方向。在与物理设备或与用户实体世界有交互的应用中,准确性与有效性将决定一个边缘智能应用是否可用。这还将包括对物理安全性、实物可行性,以及高可靠性等诸多方面的要求。无论在哪个方面,如果可能实现非常明显的进展都将有效地推动该领域的发展。

同时,随着准确性和有效性在边缘智能应用上的提高,也必将产生许多新型应用,包括与人类健康相关联的、与养老和教育相关联的,以及对人类世界自动化有更多贡献的诸多未来应用。

#### 4.5 区块链技术的整合

自比特币的创造、发展以及之后所面临的争议来看,其背后支撑的区块链技术逐步走向台前,并从其他多个方向越来越广泛地影响了诸多领域和应用。作为一个无限冗余的分布式数据库系统,区块链给学术界和工业界带来了诸多优势,同时本身存在着一些技术限制,这些都将带来机遇与挑战。首先,在边缘计算有关的智能应用中,点对点的通信与数据共识的达成将成为一个必要条件,而区块链所提供的相应技术集合有效地解决了这一系列问题。其次,区块链技术所带来的智能合约等衍生技术,可以有效地帮助边缘计算中的智能应用,将事件的触发机制上升到一个新的高

度。而且,对于追踪溯源等有关的智能应用,区块链技术将带来前所未有的便利性和可靠性。除此之外,零知识证明等安全技术也由于在区块链中的广泛应用取得了诸多进展,这一技术必将作为解决数据相互信任和隐私安全保护的重要手段而越来越多地被应用于边缘计算中数据驱动的智能应用中。

#### 5 结束语

文中,我们针对边缘计算中数据驱动的智能应用这一子领域进行了全面的综述和比较:首先从物联网的发展以及边缘计算概念出发,综合分析了边缘计算产生的必要性和重要性,并与云计算和大数据等最新发展的技术进行了对比,进一步说明边缘计算的自身特点。在将边缘计算和人工智能相结合的基础上,创新性提出面向智能应用的边缘体系架构,并针对架构中的每一部分,总结了最新文献中的研究成果。我们还对目前已经存在的智能应用,例如无人驾驶汽车、智能数据库以及金融领域智能化应用的最新成果进行了总结。最后,探讨了边缘计算中数据驱动智能应用的发展与方向、机遇和挑战。

#### 参考文献

[1] 张佳乐,赵彦超,陈兵,等.边缘计算数据安全与隐私保护研究综述[J].通信学报, 2018, 39(3):1-21. DOI: 10.11959/j.issn.1000-436x.2018037

[2] WANG S, ZHANG X, ZHANG Y, et al. A Survey on Mobile Edge Networks: Convergence of Computing, Caching and Communications [J]. IEEE Access, 2017, (5):6757-6779. DOI: 10.1109/ACCESS.2017.2685434

[3] MACH P, BECVAR Z. Mobile Edge Computing: A Survey on Architecture and Computation Offloading [J]. IEEE Communications Surveys & Tutorials, 2017, 19(3): 1628-1656. DOI: 10.1109/COMST.2017.2682318

[4] BIGELOW S J. What's Happening in Data

Center Energy Management [EB/OL]. (2014-06)[2018-12-20]. <https://searchdatacenter.techtarget.com/feature/Whats-happening-in-data-center-energy-management>

[5] VERBELEN T, SIMONENS P, TURCK F D, et al. Leveraging Cloudlets for Immersive Collaborative Applications [J]. IEEE Pervasive Computing, 2013, 12(4):30-38. DOI: 10.1109/MPRV.2013.66

[6] JABER Z Q, YOUNIS M I. Design and Implementation of Real Time Face Recognition System (RTFRS) [J]. International Journal of Computer Applications, 2014, 94(12):15-22

[7] SLEIGHT D A. Intelligent Databases: Easing Access to Information (Spring, 1993) [EB/OL]. [2018-12-04]. <https://msu.edu/~sleight/Inteldat.html>

[8] BARTELS A, PARKER A, DALEY E. Smart Computing Drives the New Era of IT Growth [EB/OL]. [2018-12-04]. <https://www.forrester.com/report/Smart%20Computing%20Drives%20The%20New%20Era%20Of%20IT%20Growth/-/E-RES55157?src=55590pdf&docid=55157>

[9] TANG B, CHEN Z, GERALD, et al. A Hierarchical Distributed Fog Computing Architecture for Big Data Analysis in Smart Cities[C]// Proceedings of the ACM BigData & Social Informatics 2015. USA: ACM, 2015. DOI: 10.1145/2818869.2818898

[10] Insights Team. Anatomy of A Smart City: Four Stars On The Global Stage [EB/OL]. (2018-10-24)[2018-12-20]. <https://www.forbes.com/sites/insights-inteliot/2018/10/24/anatomy-of-a-smart-city-four-stars-on-the-global-stage/#d6d115f52437>

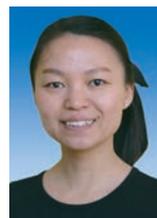
#### 作者简介



**秦永彬**, 贵州大学计算机科学与技术学院副院长、教授、博士生导师, 贵州省公共大数据重点实验室副主任, 贵州省软件工程与信息安全特色重点实验室常务副主任; 主要研究方向为机器学习、文本分析与算法设计。



**韩蒙**, 美国肯尼索州立大学(原美国南方州立理工大学)助理教授、数据驱动智能研究实验室主任; 主要研究方向为数据驱动的智能算法及深度学习, 以及与区块链技术有关的数据安全与隐私保护等。



**杨清亮**, 美国肯尼索州立大学在读硕士研究生、数据驱动智能研究实验室助理; 主要研究方向为金融科技以及与数据分析及有关的解决方案设计; 已发表会议论文2篇。



# 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 边缘计算的网路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。当数据仅在本地产有价值时,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、接口类型等典型选择与制定。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,则要按需进行资源规划,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。



# B5G 毫米波和太赫兹技术的背景、应用和挑战

## Technical Background, Applications and Challenges of B5G Millimeter Wave and Terahertz Communication

彭琳/PENG Lin<sup>1,2</sup>  
段亚娟/DUAN Yajuan<sup>1</sup>  
别业楠/BIE Yenan<sup>1,2</sup>

(1. 中兴通讯股份有限公司, 广东 深圳 518057;  
2. 移动网络和移动多媒体技术国家重点实验室,  
广东 深圳 518057)

(1. ZTE Corporation, Shenzhen 518057, China;  
2. State Key Laboratory of Mobile Network and  
Mobile Multimedia Technology, Shenzhen  
518057, China)

DOI: 10.12142/ZTETJ.201903012

网络出版地址: <http://kns.cnki.net/kcms/detail/34.1228.TN.20190614.0941.003.html>

网络出版日期: 2019-06-14  
收稿日期: 2019-05-20

**摘要:** 认为在后 5G (B5G) 和 6G 时代, 毫米波和太赫兹通信在地面无线移动通信及空间通信中都将发挥重要作用。从发展的预期来看, B5G 将扩展到毫米波高端频谱, 而 6G 将扩展到太赫兹频段。要将高频毫米波和太赫兹通信推向实际部署, 还需要解决射频器件、天线、信号处理复杂度、空间信道建模、组网的一系列关键问题。

**关键词:** 毫米波通信; 太赫兹通信; 未来移动通信; B5G

**Abstract:** In Beyond 5G (B5G) and 6G era, the millimeter wave and terahertz communication will play an important role in wireless mobile communication and space communication. B5G will extend to the high-end millimeter wave spectrum and 6G will extend to terahertz frequencies. To push the millimeter wave and terahertz communication to actual deployment, it still need to face many key challenges in radio frequency (RF) devices, antenna, signal processing complexity, space channel modeling and networking.

**Key words:** millimeter wave communication; terahertz communication; future mobile communication; B5G

### 1 技术背景

随着无线通信的快速发展, 频率在 30 GHz 以下的无线电波已经得到了广泛应用; 但是各种信息终端的大量涌现, 使得无线数据速率每 18 个月就翻一番<sup>[1]</sup>, 无线通信系统呈现出超高数据速率的爆炸性增长趋势。虽然 5G 接入的峰值速率为 1~20 Gbit/s, 但仍不能满足未来无线通信中日益增长的数据流量要求。例如, 虚拟现实 (VR) 设备的最小数据速率将达到 10 Gbit/s。

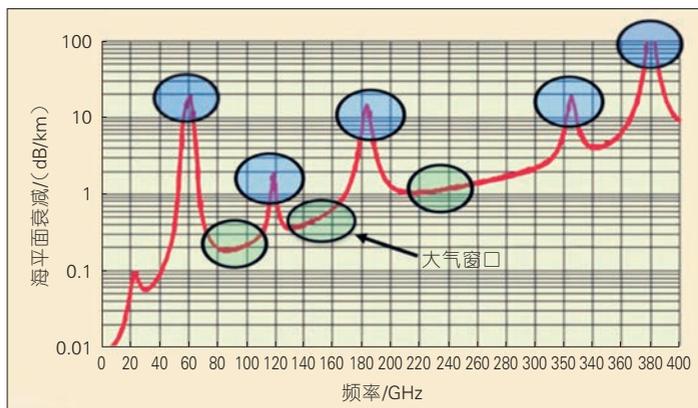
此外, 无压缩的超高清视频和 3D 视频的数据速率将分别达到 24 Gbit/s 和 100 Gbit/s<sup>[2]</sup>。要达到更高的信息传输的数据速率, 根据香农定理, 在当前常规频段带宽受限条件, 可以通过努力进行改进。但如果要实现数据传输速率达到 100 Gbit/s 甚至 1 Tbit/s, 改进仍然是不切实际或者代价巨大的。为了更好地解决这一问题, 需要探索新的频段。

毫米波和太赫兹波, 是一种介于无线电与光之间的电磁波。从 30 GHz 到 3 THz 的频率仍然没有得

到充分利用。毫米波和亚太赫兹频段的大气衰减特性如图 1 所示<sup>[3]</sup>, 存在一些宽频带的“大气窗口”, 具有较低的大气衰减。如果能保证这个范围内的宽频带, 就有可能使用高阶调制以超过 100 Gbit/s 的速度进行高速无线通信<sup>[3]</sup>。随着 5G 的逐步商用落地, 业界也将越来越多关注毫米波和太赫兹频段。

### 2 技术发展

5G 定义了增强移动宽带 (eMBB)、超可靠低时延 (uRLLC)、



▲图1 大气吸收特性随频率的变化

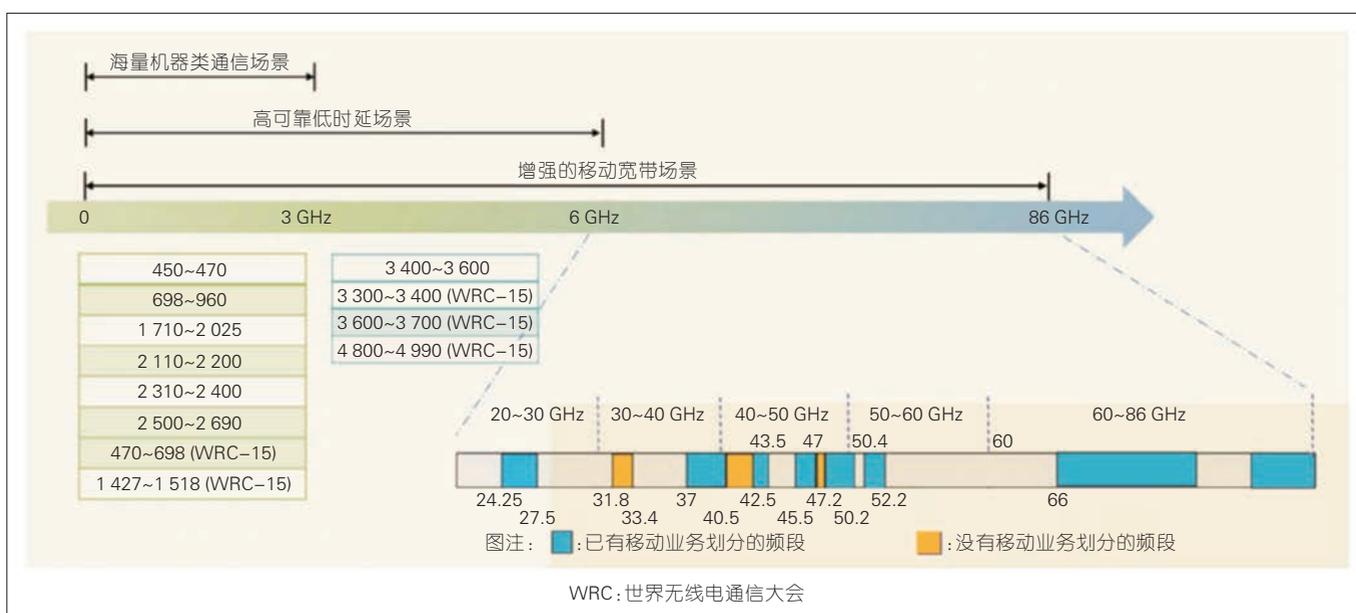
海量机器类通信(mMTC)3大场景。其中eMBB场景针对速率的提升,主要应对集城区、室内热点等大流量场景;uRLLC场景针对网络时延的降低,主要应用于无人驾驶、智能工厂等低时延场景;mMTC场景针对海量大连接,对应的是物联网等海量机器类通信的场景。5G频率涵盖低、中、高频段:6 GHz以下的中、低频段覆盖能力强,能够实现全网覆盖,满足uRLLC场景和mMTC通信场景;6 GHz以上的高频段拥有连续大带宽,满足eMBB场

景,如图2<sup>[5]</sup>所示。

全球5G频率规划工作主要在国际电信联盟(ITU)、第三代合作伙伴计划(3GPP)等国际标准化组织的框架下开展。面对2020年后国际移动通信(IMT)系统对频谱资源需求的激增,2019年即将举行的世界无线电通信大会(WRC-19)专设1.13议题,在24.25~86 GHz频率范围的若干个候选频段中为5G寻找新增频段,并在ITU-R第5研究组特设了TG 5/1工作组专门负责该议题的研究。目前最可能优先

部署26 GHz/28 GHz/39 GHz,同时随着3GPP R15定义eMBB与固定无线接入(FWA)业务后,毫米波产业链也迅速被催熟。3GPP R17还将研究52.6~114.25 GHz频段。

针对下一代通信演进,100 GHz~3 THz之间的太赫兹频段的研究也引起越来越多的关注。日本也将开发太赫兹技术列为“国家支柱技术十大重点战略目标”之首。日本NTT早在2006年在国际上首次研制出0.12 THz无线通信样机,并于2008年北京奥运会上成功用于高清转播<sup>[6]</sup>,目前正在全力研究0.5~0.6 THz高速大容量无线通信系统。日本总务省规划将在2020年东京奥运会上采用太赫兹通信系统实现100 Gbit/s高速无线局域网服务<sup>[4]</sup>。欧盟已将发展太赫兹通信列为了6G研究计划。ITU在WRC-19大会专设议题1.15,以确定运行在275 GHz以上频段的陆地移动和固定业务系统的技术与操作特性,



▲图2 5G关键场景对应频谱

包括研究相关频谱需求、建立 0.275~0.475 THz 频段范围内的传播模型、开展业务间电磁兼容分析、确定候选频段等。美国联邦通信委员会(FCC)于2019年3月正式开放面向未来6G网络服务的“太赫兹频谱”,实验频谱范围为 95 GHz~3 THz。

### 3 应用场景

#### 3.1 无线移动通信

高频毫米波、太赫兹频段有望在不久的将来实现几十吉比特每秒甚至太比特每秒的数据速率。从无线通信的发展趋势来看,高频毫米波、太赫兹无线通信系统具有良好的应用前景,具体的覆盖场景如图3所示。

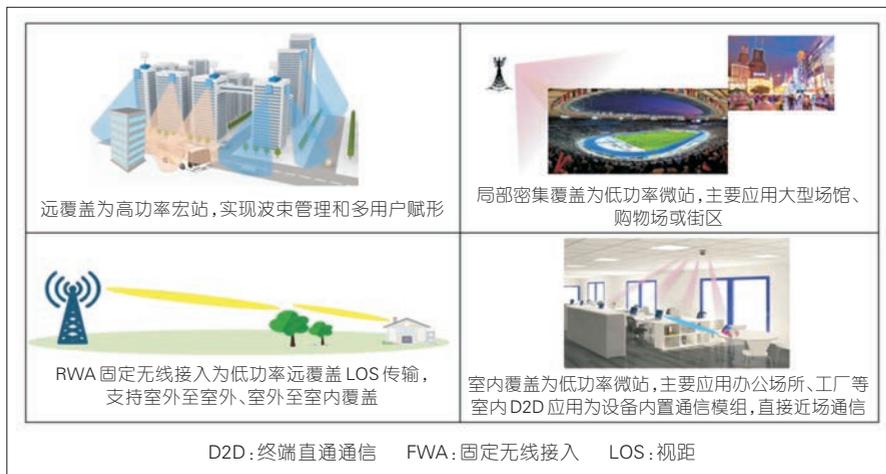
目前,中兴通讯 26 GHz 毫米波基站有源天线单元(AAU)产品已经全面支持上述场景;但到了更高的毫米波、太赫兹波段,由于传输损耗及水汽影响和当前器件输出功率发展的限制,我们认为高频毫米波、太赫兹无线通信的发展需要经过先

提升速率后提升覆盖距离的过程。最先可能广泛应用于室内应用场景,包括室内蜂窝网络、无线局域网(WLAN)、无线个域网(WPAN)<sup>[7]</sup>。对于WPAN系统有望实现手机、笔记本电脑、耳机等桌面设备的特定应用。对于室内蜂窝网和WLAN系统,我们认为可以在高/地铁站、机场、办公场所等人流密集的开阔型室内场所部署。与采用低频段室内覆盖不同,高频毫米波、太赫兹天线能够用窄波束同时向不同方向的多用户传输信息。超高的数据速率和超低延时技术将支持用户在室内体验高品质的视频服务。尤其目前虚拟现实(VR)技术的发展受到低数据速度无线通信的严重限制,毫米波、太赫兹带宽应用于无线通信系统后,无线VR技术将带来比有线VR系统更好的用户体验,将推动各种现实(XR)技术进一步快速发展。

#### 3.2 空间通信网络

随着通信技术的演进,无所不在的网络是未来网络的重要特征,

其中一个重要途径是对空间通信网络的发展以及与地面网络的融合。为了满足空间通信网络的发展需求,新的频谱需要提供极高的数据速率传输。毫米波、太赫兹信号的高大气衰减大大缩短了地面通信系统的通信距离和传输速率;但与地面上的毫米波、太赫兹通信相反,在无大气环境中的空间应用则可不受大气衰减的影响,这对毫米波、太赫兹波段的空通信是非常重要的。基于毫米波、太赫兹的卫星通信可用于星地间骨干链路、星间骨干链路、星-浮空平台间链路、星-飞行器间链路、飞行器/浮空平台与地面间链路,实现大容量信息传输<sup>[8]</sup>。此外,还有一个太赫兹的特殊应用例子:当高速飞行器飞进大气层后,由于激波产生高温使空气电离,并形成等离子体鞘包裹在飞行器外部。通常等离子体鞘频率在 60~70 GHz 左右,传统的测量和通信方法难以穿透等离子体鞘层。然而,太赫兹波频率远高于等离子体鞘层频率,可以穿透等离子体鞘层对飞行器进行通信和测量。



▲图3 毫米波、太赫兹无线覆盖场景

### 4 关键技术挑战

#### 4.1 毫米波/太赫兹射频器件

在未来的毫米波、太赫兹波段通信中,射频器件在硬件电路方面存在一些挑战。

(1)器件建模。随着半导体器件工作频率达到太赫兹频段,半导体材料的影响和器件封装在高频频段的分布参数效应越来越突出。非线性模型的交流和半导体器件的参

数提取直接影响电路设计和系统性能的确立;因此,利用一些新颖的设计方法来获得半导体器件的精确参数对硬件电路设计具有重要意义。

(2)太赫兹调制器。作为太赫兹电路的核心部件,根据电路实现方式可分为混频调制和直接调制,其中混频调制器可以支持高阶调制方式,可以大大提高频率谱利用率,实现较高的传输速率,但缺点是目前本振倍频相噪恶化和变频损耗大;直接调制器能实现太赫兹波幅度和相位的直接调制,可灵活搭配中高功率太赫兹源,是实现中远距离无线通信的有效途径,已成为近年来研究的热点,但目前直接调制器的调制速率和深度还有待提高。

(3)毫米波/太赫兹系统芯片集成的挑战。在高频毫米波、太赫兹频段,由于不同半导体工艺特征频率( $f_t$ )的存在不同如表1,要融合多种工艺(如硅基工艺和III-V基工艺等)各自独特优势,实现高集成度、高性能、多功能的单片电路,异质集成电路技术的研究具有重要意义。近期随着硅基工艺能力不断提升,截止频率也在提升;目前看功率放大器(PA)适合用氮化镓(GaN)工艺,RX适合用磷化铟(InP)工艺,单片集成电路可以用异质集成方式来

实现。

(4)太赫兹芯片封装的挑战。由于电感电抗的存在,随着工作频率的增加,传统的键合线通过射频信号变得损耗难以接受;因此,必须开发一种用于射频信号输入和输出的低损耗方法。

#### 4.2 毫米波/太赫兹天线

高频毫米波、太赫兹频段,为了补偿比低频更加明显的大气衰减,需要采用高增益天线来补偿。虽然在高精度机械加工的基础上,应用传统的卡塞格伦天线和抛物面天线,可以实现60 dB以上的高增益天线。然而,B5G/6G移动通信系统的集成度和小型化是重要的考量,需要更小的微型天线,如平面贴片天线阵。在毫米波、太赫兹移动通信应用中,需要快速波束扫描、跟踪、切换等波束管理和多用户赋形,这类天线应具有高增益和快速的波束扫描功能,波束成型芯片和阵列天线是关键器件。因此,研究高性能的平面多功能天线阵列毫米波系统,支持不同的应用场景应用,是一项十分重要的工作。

#### 4.3 超宽带低复杂度信号处理

在B5G/6G时代中,毫米波、太

赫兹通信有望使得100 Gbit/s甚至1 Tbit/s的高速通信成为可能,这要求模拟数字转换器(ADC)的采样率随着通信带宽的增大而增大。然而,要制造出满足低功耗、小尺寸和高带宽要求的ADC越来越困难。例如,高速(如 $\geq 10$  G Sample/s)和高精度(如 $\geq 12$ 位)ADC实现难度大,过于昂贵且功耗大。为了解决这一问题,采用低分辨率ADC进行量化和ADC时间交织技术都是值得研究的方向。

#### 4.4 空间信道模型

在无线通信系统中,建立有效的信道模型对最大化带宽分配,提高频谱效率至关重要。常见的太赫兹波传播模型可以由视距(LOS)和非视距(NLOS)2种波传播模型组成。太赫兹信道的移动性,至少对于未来的应用来说,仍然是一个值得进一步探索的关键挑战。由于混合波束形成结构和大量天线的存在,太赫兹通信系统的信道估计面临着很大的挑战。由于用户的移动性往往导致太赫兹波信道变化较快,基站(BS)获取准确的太赫兹波信道信息也具有一定的挑战性。因此,需要更有效的信道估计方案来利用时变信道的时变相关性。

#### 4.5 网络组网架构和空口的优化

未来,随着高频毫米波、太赫兹技术在不同通信场景中的广泛应用,毫米波、太赫兹环境下的无线通信组网必将成为研究热点。根据目前的研究,与低频段相比,高频毫米波、太赫兹频段的波束相对较窄,波

▼表1 不同半导体工艺对比

| 工艺   | CMOS              | SiGe      | GaAs  | GaN                   | InP              |
|------|-------------------|-----------|---|-----------------------|------------------|
| 特征频率 | <200 GHz          | <200 GHz  | <500 GHz  | <200 GHz              | 200~500 GHz      |
| 实现能力 | CMOS的功率可以到SiGe的功率 | 小于GaAs的功率 | 目前功率已经到极限,因为工艺耐压能力有限,提升功率必须高压容易击穿;NF性能在THz频段无法提升;Mixer的变频损耗有天然缺陷; | 功率是GaAs的5倍,成熟度不够,缺陷较多 | 功率不高,不耐压,但是NF性能好 |

CMOS: 互补金属氧化物半导体  
GaAs: 砷化镓

GaN: 氮化镓  
InP: 磷化铟

NF: 噪声系数  
SiGe: 锗硅

束的覆盖范围有限,且性能易受障碍物、路损,以及天气变化影响;因此,由全向天线实现的全向组网技术不能满足快速查找网络节点、成全向组网和节约系统能耗的要求。网络组网架构需要进一步优化,需要进一步研究无线网状网络(MESH)与多跳组网、点到多点、终端合作、灵活双工与接入、分布式或网络多输入多输出(MIMO)及人工智能(AI)网优等新的网络拓扑与结构优化技术来提升 B5G 的网络性能、效率与灵活性。

#### 4.6 空口与高频段组合技术

(1) 频谱灵活使用与管理。毫米波/太赫兹大带宽(例如 3~5 GHz)频谱的许可、使用、管理与经营方式相对传统低频段频谱将可能发生很大的变化。这部分频谱是独占式的许可频谱,非独占方式的非许可、频谱接入共享,还是地区或应用特定的微许可,以及将蜂窝无线接入与前传/回传频谱共享的 IAB 方式,都将对技术带来挑战。

(2) 基带传输与空口设计。具体包括与毫米波/太赫兹信道传播模型及组网拓扑相匹配的波形、帧结构、参数集、信道编码、高阶调制、自适应链路、超窄波束管理、灵活双工模式、干扰管理与消除、调度、接入、移动性管理、覆盖扩展等方案设计,尤其是引入 AI/机器学习(ML)技术之后的空口设计与优化技术。

(3) 终端复杂性与能耗降低。如前所述,毫米波/太赫兹芯片与超大带宽信号处理对网络侧站点或客户终端设备(CPE),以及终端侧的

复杂性和能耗(包括电池)都有很大的影响;因此,从低频段向高频段的演进,始终需要进一步降低终端的复杂性与能耗,以满足网络在单位焦耳能耗上实现的数据比特率等性能和终端耗电与充电等要求。

## 5 结束语

根据业界对 B5G/6G 演进的预期,更高频毫米波和太赫兹无线通信系统已成为全球研究的热点。拥有大带宽的高频毫米波和太赫兹频段是面向未来通信的频谱宝库,其中太赫兹技术更可能是 6G 时代的重要技术组成部分。毫米波和太赫兹通信不仅能够面向未来超高速率、超低时延无线移动通信网络应用,实现未来的全频谱接入,还可支持空天一体化的全覆盖需求。标准组织、运营商和设备厂家以及产业链各方都在积极努力,以实现毫米波和太赫兹无线通信的早日成熟。虽然由于更高毫米波、太赫兹通信的发展周期较短,需要进一步发展来探索毫米波、太赫兹频段的特性和高性能器件,但我们相信该技术必将成为未来超高速无线通信的重要组成部分。

### 致谢

本文得到了中兴通讯无线产品经营部方敏、李师波、段向阳、段斌、王永贵、郁光辉、胡留军的鼎力帮助,谨致谢意!

### 参考文献

- [1] CHERRY S. Edholm's Law of Bandwidth [J]. IEEE Spectrum, 2004, 41(7): 58. DOI:10.1109/mspec.2004.1309810

- [2] 陈智,张雅鑫,李少谦. 发展中国太赫兹高速通信技术与应用的思考[J]. 中兴通讯技术, 2018, 24(3): 43-47. DOI:10.3969/j.issn.1009-6868.2018.03.008
- [3] HARA S, TAKANO K, KATAYAMA K, et al. 300-GHz CMOS Transceiver for Terahertz Wireless Communication[C]//2018 Asia-Pacific Microwave Conference (APMC) USA: IEEE, 2018: 429-431. DOI:10.23919/apmc.2018.8617468
- [4] RAPPAPORT T S, MURDOCK J N, GUTIERREZ F. State of the Art in 60-GHz Integrated Circuits and Systems for Wireless Communications [J]. Proceedings of the IEEE, 2011, 99(8): 1390. DOI:10.1109/jproc.2011.2143650
- [5] 陈佳佳,王坦,曾昱祺. 全球 5G 频谱发展动态[J]. 电信技术, 2019(1): 13
- [6] HIRATA A, HARADA M, NAGATSUMA T. 120-GHz Wireless Link Using Photonic Techniques for Generation, Modulation, and Emission of Millimeter-Wave Signals [J]. Journal of Lightwave Technology, 2003, 21(10): 2145. DOI:10.1109/jlt.2003.814395
- [7] SONG H J, NAGATSUMA T. Present and Future of Terahertz Communications [J]. IEEE Transactions on Terahertz Science and Technology, 2011, 1(1): 256. DOI:10.1109/tthz.2011.2159552
- [8] NAGATSUMA T, KASAMATSU A. Terahertz Communications for Space Applications[C]//2018 Asia-Pacific Microwave Conference (APMC). New York, USA: IEEE, 2018: 73-75. DOI:10.23919/apmc.2018.8617598

### 作者简介



明专利 3 项。

彭琳,中兴通讯股份有限公司 B5G/6G 射频技术预研负责人、项目经理,负责 B5G/6G 射频技术预研论证、关键技术规划等工作;带领团队开发多款 WiMax、LTE 宏站 RRU 机型及 Qcell 室内覆盖 RRU 机型;负责研发的 Qcell 数字室内覆盖方案曾获 LTE Awards 创新大奖;获国家发明



段亚娟,中兴通讯股份有限公司 RRU 射频 TRX 专业团队负责人;负责 TRX 器件规划、定义,以及 TRX 链路方案设计和器件导入工作;带领团队规划和定义了多款核心芯片,使中兴通讯射频部分链路竞争力一直保持业界领先;获国家发明专利 7 项。



别业楠,中兴通讯股份有限公司无线研究院副院长、硬件研发中心主任;拥有超过 20 年的电信行业研发管理经验;负责研发的业界第一款高效率、超宽带 RRU 曾获得了深圳市科技进步一等奖;发表论文 1 篇,获国家发明专利 8 项。

# 《中兴通讯技术》杂志(双月刊)投稿须知

## 一、杂志定位

《中兴通讯技术》杂志为通信技术类学术期刊。通过介绍、探讨通信热点技术,以展现通信技术最新发展动态,并促进产学研合作,发掘和培养优秀人才,为振兴民族通信产业做贡献。

## 二、稿件基本要求

### 1. 投稿约定

- (1)作者需登录《中兴通讯技术》投稿平台:tech.zte.com.cn/submission,并上传稿件。第一次投稿需完成新用户注册。
- (2)编辑部将按照审稿流程聘请专家审稿,并根据审稿意见,公平、公正地录用稿件。审稿过程需要1个月左右。

### 2. 内容和格式要求

- (1)稿件须具有创新性、学术性、规范性和可读性。
- (2)稿件需采用WORD文档格式。
- (3)稿件篇幅一般不超过6000字(包括文、图),内容包括:中、英文题名,作者姓名及汉语拼音,作者中、英文单位,中文摘要、关键词(3~8个),英文摘要、关键词,正文,参考文献,作者简介。
- (4)中文题名一般不超过20个汉字,中、英文题名含义应一致。
- (5)摘要尽量写成报道性摘要,包括研究的目的、方法、结果/结论,以150~200字为宜。摘要应具有独立性和自明性。中英文摘要应一致。
- (6)文稿中的量和单位应符合国家标准。外文字母的正斜体、大小写等须写清楚,上下角的字母、数据和符号的位置皆应明显区别。
- (7)图、表力求少而精(以8幅为上限),应随文出现,切忌与文字重复。图、表应保持自明性,图中缩略词和英文均要在图中加中文解释。表应采用三线表,表中缩略词和英文均要在表内加中文解释。
- (8)所有文献必须在正文中引用,文献序号按其在文中出现的先后次序编排。常用参考文献的书写格式为:
  - 期刊[序号]作者.题名[J].刊名,出版年,卷号(期号):引文页码.数字对象唯一标识符
  - 书籍[序号]作者.书名[M].出版地:出版者,出版年:引文页码.数字对象唯一标识符
  - 论文集中析出文献[序号]作者.题名[C]//论文集编者.论文集名(会议名).出版地:出版者,出版年(开会年):引文页码.数字对象唯一标识符
  - 学位论文[序号]作者.题名[D].学位授予单位所在城市名:学位授予单位,授予年份.数字对象唯一标识符
  - 专利[序号]专利所有者.专利题名:专利号[P].出版日期.数字对象唯一标识符
  - 国际、国家标准[序号]标准名称:标准编号[S].出版地:出版者,出版年.数字对象唯一标识符
- (9)作者超过3人时,可以感谢形式在文中提及。作者简介包括:姓名、工作单位、职务或职称、学历、毕业于何校、现从事的工作、专业特长、科研成果、已发表的论文数量等。
- (10)提供正面、免冠、彩色标准照片一张,最好采用JPG格式(文件大小超过100kB)。
- (11)应标注出研究课题的资助基金或资助项目名称及编号。
- (12)提供联系方式,如:通讯地址、电话(含手机)、Email等。

### 3. 其他事项

- (1)请勿一稿两投。凡在2个月(自来稿之日算起)以内未接到录用通知者,可致电编辑部询问。
- (2)为了促进信息传播,加强学术交流,在论文发表后,本刊享有文章的转摘权(包括英文版、电子版、网络版)。作者获得的稿费包括转摘酬金。如作者不同意转摘,请在投稿时说明。

编辑部地址:安徽省合肥市金寨路329号凯旋大厦1201室,邮政编码:230061

联系电话:0551-65533356,联系邮箱:magazine@zte.com.cn

本刊只接受在线投稿,欢迎访问本刊投稿平台:tech.zte.com.cn/submission

# 中兴通讯技术

ZTE TECHNOLOGY JOURNAL

办刊宗旨:

以人为本,荟萃通信技术领域精英  
迎接挑战,把握世界通信技术动态  
立即行动,求解通信发展疑难课题  
励精图治,促进民族信息产业崛起

---

双月刊 1995年创刊 总第146期  
2019年6月 第25卷 第3期

主管:安徽出版集团有限责任公司  
主办:时代出版传媒股份有限公司  
深圳航天广宇工业有限公司  
出版:安徽科学技术出版社  
编辑、发行:中兴通讯技术杂志社

总编:王喜瑜  
副总编:蒋贤骏  
常务副总编:黄新明  
责任编辑:徐焯  
编辑:卢丹、朱莉  
排版制作:余刚  
发行:王萍萍  
编务:王坤

---

《中兴通讯技术》编辑部  
地址:合肥市金寨路329号凯旋大厦1201室  
邮编:230061  
网址:tech.zte.com.cn  
投稿平台:tech.zte.com.cn/submission  
电子信箱:magazine@zte.com.cn  
电话:(0551)65533356

传真:(0551)65850139  
发行范围:公开发行  
印刷:合肥添彩包装有限公司  
出版日期:2019年6月20日  
中国标准连续出版物号:ISSN 1009-6868  
CN 34-1228/TN  
定价:每册20.00元