

工业边缘计算研究现状与展望

王其朝^{1, 2, 3, 4}, 金光淑⁵, 李 庆^{1, 2, 3, 4}, 王 锴^{1, 2, 3}, 杨祖业⁵, 王 宏^{1, 2, 3}

1. 中国科学院网络化控制系统重点实验室, 辽宁 沈阳 110016; 2. 中国科学院沈阳自动化研究所, 辽宁 沈阳 110016;
3. 中国科学院机器人与智能制造创新研究院, 辽宁 沈阳 110169; 4. 中国科学院大学, 北京 100049;
5. 沈阳中科博微科技股份有限公司, 辽宁 沈阳 110179

基金项目: 国家重点研发计划资助项目(2020YFB1709702); 国家自然科学基金资助项目(62073313)

通信作者: 王宏, wang@sia.cn 收稿/录用/修回: 2021-02-09/2021-05-08/2021-05-24

摘要

随着工业互联网技术的发展, 工业设备生成的数据量以指数形式增长, 给云计算带来了巨大的数据压力. 边缘计算由于低时延、低流量、强隐私性等特点在工业领域中有了越来越多的应用研究. 本文针对近几年的工业云计算和边缘计算技术进行了综述. 首先介绍了工业互联网背景下云计算和边缘计算的发展历史, 分析了边缘计算的定义和边缘计算的几种计算典型形式与工业边缘计算系统的关系. 其次分析了工业边缘计算的几种典型应用, 针对当前的研究现状, 讨论了工业场景中影响边缘计算应用的几种关键技术, 并对工业场景下的研究挑战做了总结与展望.

关键词

边缘计算
云计算
工业互联网
智能制造
计算虚拟化
中图法分类号: TP18
文献标识码: A

Industrial Edge Computing: Vision and Challenges

WANG Qizhao^{1, 2, 3, 4}, JIN Guangshu⁵, LI Qing^{1, 2, 3, 4}, WANG Kai^{1, 2, 3}, YANG Zuye⁵, WANG Hong^{1, 2, 3}

1. Key Laboratory of Networked Control Systems, Chinese Academy of Sciences, Shenyang 110016, China;
2. Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China;
3. Institutes for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences, Shenyang 110169, China;
4. University of Chinese Academy of Sciences, Beijing 100049, China;
5. Microcyber Corporation, Shenyang 110179, China

Abstract

With the development of industrial Internet technology, the data volume of industrial equipment has increased exponentially, which has brought a huge data pressure to cloud computing. There is more and more applications and researches of edge computing in the industrial field due to the characteristics of low latency, low traffic, and good privacy. We review the research on cloud computing and edge computing in industrial applications in recent years. We firstly introduce the development history of cloud computing and edge computing in the context of the industrial Internet, and analyze the definition of edge computing and the relationships between several typical forms of edge computing and industrial edge computing system. Secondly, several typical applications of industrial edge computing are analyzed. In view of the current research status, several key technologies affecting edge computing applications in industrial scenarios are discussed. Finally, the research challenges in the industrial scene are summarized and prospected.

Keywords

edge computing;
cloud computing;
industrial Internet;
smart manufacturing;
computing virtualization

0 引言

中国是一个制造业大国, 智能制造作为中国制造 2025 的主攻方向, 对中国从制造大国到制造强国的转变有着重要的意义^[1]. 相比于传统制造业, 智能制造需要实现制造过程的数字化、网络化和智能化, 加快信息技术与制造技

术的融合发展, 全面提升研发、生产、管理和服务的智能化. 工业互联网作为智能制造的基础设施, 是连接机器、物料、人和信息系统的基础网络, 建设低时延、高可靠、广覆盖的工业互联网是实现智能制造的基本保障. 数据是信息是载体, 获取有价值的信息往往需要对大量数据进行计算分析. 数据的计算分析需要强大的计算资源来支持,

从传统的互联网发展经验来看,这种计算资源由统一的服务商作为一种服务提供是一种很好的解决方案,也就是云计算^[2]。工业互联网中,制造商通过工业云平台收集大量不同地域分布设备的生产数据,利用数据分析技术如深度学习进行数据挖掘,获取有价值的信息。工业云平台技术继承自传统的云计算,云计算服务通常由传统的云计算服务提供商提供,针对工业需求的特点提供一些软件及服务(SaaS)、平台及服务(PaaS)及基础设施及服务(IaaS)等工业云服务。工业云计算与传统的云计算在技术上没有本质的区别,包括采用容器或者虚拟机作为虚拟化技术^[3-4],通过KVM(Kernel-based virtual machine)、DockerSwarm及Kubernetes等^[5-7]技术对虚拟机或容器进行管理。

工业云计算中数据主要来源为工业现场,这需要分布在工业现场的计算设备支持,它们收集生产数据并发送至云端。单纯的云计算由于时延不定、隐私泄露和通信成本高等问题,无法满足工业环境中的计算需求,因此需要利用工业现场的计算设备辅助完成云端计算,这些位于工业现场的计算设备构成了工业互联网中的边缘计算^[8]。边缘计算是近几年提出的新兴概念,主要思想是将计算设备下沉的数据边缘侧以补足远端云计算在通信能力方面的不足^[9]。工业现场存在的大量计算设备原生的组成了边缘计算层,这些边缘计算设备具有不同的性能、架构和任务等。如何利用这些边缘计算设备,与工业云平台协同完成相关的计算任务,是工业边缘计算的主要研究目标。

本文从边缘计算的形式与定义、工业边缘计算的典型架构及应用、工业边缘计算平台与标准化进程、边缘计算关键技术及工业应用的不足与挑战等几个方面对工业边缘计算的研究现状进行综述,以便于为相关研究和工程技术人员提供一定参考,促进边缘计算在工业领域的落地。

1 什么是边缘计算

边缘计算概念的诞生是实际需求推动的,因为云计算无法满足计算需求,从而将计算下放到边缘侧,来补充云计算的功能。边缘计算在提出之初并没有严格的定义,2016年,文[9]将边缘计算定义为:在数据源到云端之间任意位置的计算。这是一种比较宽泛的边缘计算定义,重点定义了边缘的位置,但是没有给出具体的计算形式。边缘计算从提出到现在已经成为当今研究的热点,包括成为了重要的投资领域^[10],资本的裹挟给边缘计算的研究带来了一定负面影响,使得边缘计算概念被滥用,仿佛一夜之间所有不在云端的计算都可以称之为边缘计算。尤其是在工业环境中,如工业现场的PID(Proportional-integral-derivative)控制器、显示程序、传感器采集程序,它们都是在数据产生的边缘侧进行计算,最终的结果也都可以上传到云端,如果将这些形式的计算都算作边缘计算,那么边缘计算的历史当追溯到1971年Intel公司的第一片微处理器的诞生。因此明确边缘计算定义对边缘计算的研究有着积极意义。

边缘计算是相对于云计算而提出的概念,可以理解作为一种离数据源更近的云计算。云计算最重要的特点是通过虚拟化技术将计算作为一种服务提供。边缘计算与云计算

相比,在计算位置上与云计算不同,计算形式类似。在计算位置方面,计算离云端越近,计算设备的性能越强,通信时延越高。计算形式方面,边缘计算与云计算同样可以将计算作为一种服务供其他设备使用。不同的是,边缘设备既可以贡献资源,又可以利用资源。因此,本文认为一种更狭义的边缘计算定义应该是,在数据产生的源头到云计算之间的任意位置处,有边缘设备参与的一种计算形式,边缘设备指的是可以利用特定形式(包括但不限于容器、虚拟机等计算虚拟化技术)来贡献自己或者使用其他设备计算资源的一类设备,计算资源可共享是边缘设备的最重要的特性。在这种定义下,可以很容易地区分边缘计算和传统的嵌入式计算。以智能驾驶汽车为例,如果其先进驾驶辅助系统(advanced driver assistance system, ADAS)^[11]只是在本地做感知计算,然后将结果上传到云端,那么这种形式的计算只能算作传统的嵌入式计算;如果ADAS可以通过5G、WiFi、D2D(device-to-device)等通信方式,将自己计算的一部分放到MEC服务器、微云或者其他边缘设备上,那么这样的计算形式可以认定为边缘计算。

2 边缘计算的典型形式与工业架构

边缘计算由多方不同的代表提出并主导发展,典型边缘计算形式有移动边缘计算(mobile edge computing, MEC, 也称“多接入边缘计算”, multi-access edge computing)^[12-13]、微云计算(cloudlet)^[14]和雾计算(fog computing)^[15]。

2.1 移动边缘计算

移动边缘计算的主导者是电信运营商,其动机在于电信运营商的基站是移动互联网用户接入网络的第一跳位置,在基站侧部署计算设备有利于为用户提供更为便捷的计算服务,并可以拓展其业务维度,获得更多收益。同时,5G技术的发展使得电信运营商的基站部署密度更高,而5G技术带来的高通信速度加重了核心网的通信压力,因此必须将部分核心网的功能下放到边缘侧,这也成为5G的标准之一,进一步地促进了移动边缘计算的发展^[16]。移动边缘计算中主要的边缘计算设备是位于通信基站侧的边缘服务器,边缘服务器通过核心网与云端相连,用户通过移动网络接入,后期拓展为多接入方式。与其他形式的边缘计算相比,移动边缘计算典型的特点是用户多通过移动网络接入,因此移动边缘计算的研究通常要考虑时频等移动通信资源的管理^[17]。

2.2 微云计算

微云计算的主导者为云计算服务提供商,云计算服务提供商的服务器分布在全球范围内的几个地方,不同地理位置的用户可以就近地利用云服务。用户离云服务器的地理位置越近越能享受更便捷的计算服务。而当云计算服务商进一步提高服务器的地理覆盖范围,在城市范围内提供更广泛的服务器分布时,这些服务器就形成了一个小小的微云(边缘云)。用户利用近端的微云计算可以获取更好的计算服务,云计算服务商也可以降低通信成本,阿里云^[18]、亚马逊云^[19]等都提供了类似的服务。微云计算与云计算亲缘关系最近,所有者都是云计算服务提供商,因此微云

计算研究通常需要考虑计算资源定价策略等问题^[20]。

2.3 雾计算

雾计算^[15]的主导者是以思科为代表的通信设备制造商,核心思想是利用如交换机、路由器等广泛分布的通信设备的闲置资源为用户提供更好的计算服务。雾计算的最主要特点是其计算设备的所有者是多个不同的主体,不同于移动边缘计算中计算设备属于通信运营商、微云计算中计算设备属于云计算服务提供商,雾计算中的设备没有统一的主体,计算设备的所有者拥有该设备计算和通信资源,他们通常既是资源的提供者又是资源的使用者。使用雾计算中其他计算资源需要付费,贡献自己的计算资源又可以获得收益,因此,资源的贡献激励研究往往是雾计算所关注的。

2.4 工业边缘计算的系统架构

工业场景中的边缘计算可以采用上述形式的边缘计算,但又不尽相同。移动边缘计算、微云计算和雾计算都是为了补充云计算的不足,计算是从云端下放到边缘侧,

原始的边缘侧并没有计算,而工业边缘计算却并非是这样。自云至边的发展形式,云计算的概念虽然提出很久,但是真正形成规模应用却是近几年的事情,而在工业领域的应用更是远滞后于其他领域,工业云平台的建设与边缘计算是同步进行的。这导致了工业场景中,边缘侧的计算设备是先于云计算存在的。在云计算概念没有提出前,工业现场就已经存在了大量的计算设备,并运行相关的计算任务。工业互联网的发展促进了设备联网的需求,由于工业现场总线无法直接上传数据至互联网中,因此网关开始在工业现场大量部署。工业现场的这些新旧计算设备构成了工业互联网中的边缘计算层,它具有不同的架构、资源水平、任务需求等,异构性是它们最显著的特点。因此,从边缘计算的起源来看,雾计算是与工业边缘计算最为接近的一种边缘计算形式。但是在大的工业互联网背景下,移动边缘计算、微云计算等形式的边缘计算也都可以算作工业场景下边缘计算的组成部分。因此考虑到所有边缘计算的形式,工业边缘计算的整体架构图可以如图1所示。

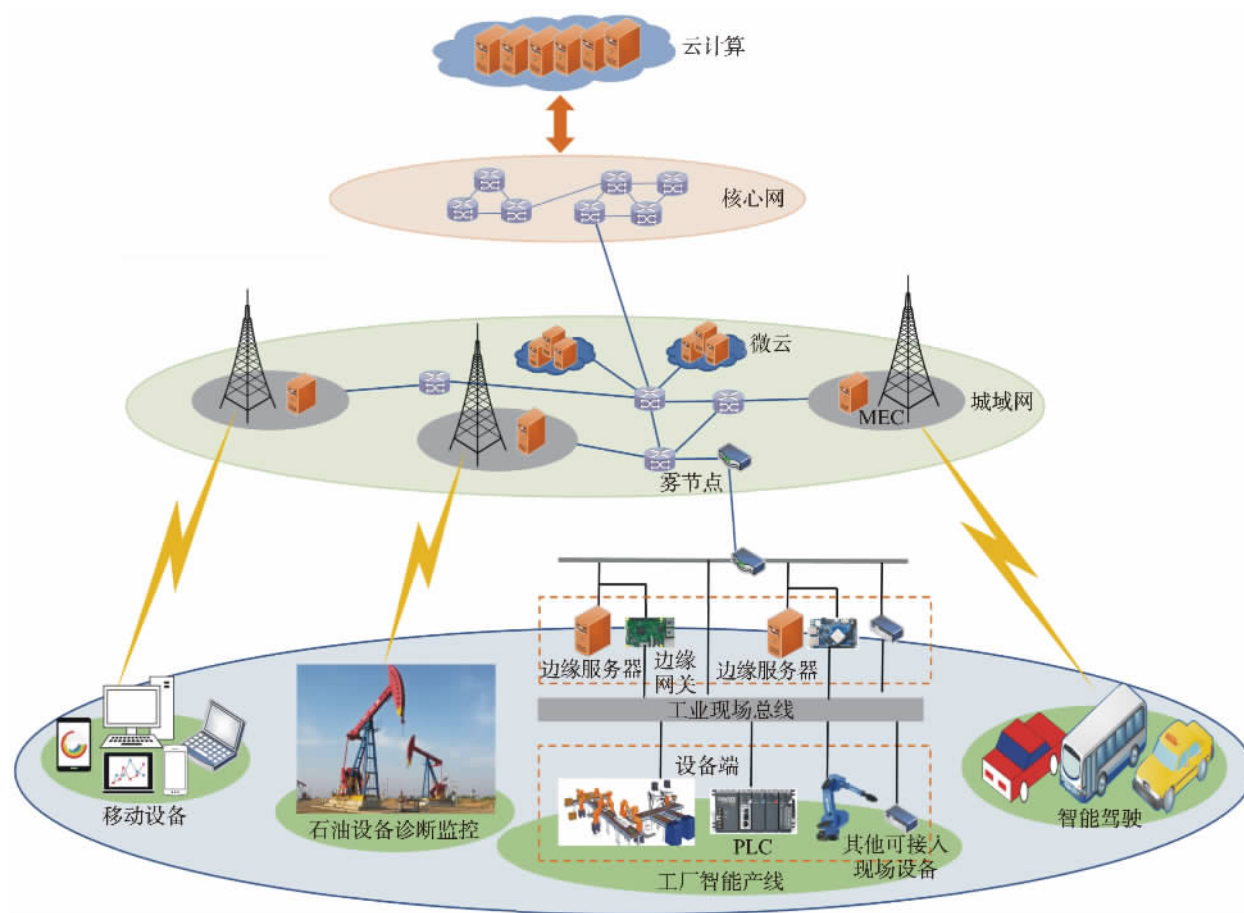


图1 工业边缘计算系统架构图

Fig.1 Industrial edge computing system architecture

该架构图可以分为云端、不同形式的边缘端和终端设备等三个部分。其中,云平台位于核心网远端,计算设备由云计算服务商提供,用户根据自己的需求选择云计算商不同层次的服务(如IaaS、PaaS和SaaS等)并部署自己的

计算任务。当云计算服务商在用户所在的城域网范围内部署了微云,用户可以利用更快速便捷的微云计算。对于通过移动网络接入的工业设备,移动运营商部署的MEC服务器也可以提供就近的计算服务。在工业现场,用户可以

利用厂房园区内的工作站进行边缘计算，也可以在现场部署边缘网关和边缘服务器，现场原有的计算设备包括各种工业控制器、网络设备等也可以成为边缘计算设备，这种形式的边缘计算与雾计算接近。以工业环境中最典型的智能装备维护为例，用户要实现收集工业现场的生产数据，并在云端进行数据分析故障诊断与参数优化等功能，首先需要利用云服务提供商提供的云计算或者微云部署云平台，对于现场数据采集、清洗和上传等功能，可以利用本地的边缘服务器、边缘网关进行。当智能装备工作在野外如石油装备等，则可以借助 MEC 的服务器。

3 边缘计算的工业应用现状

近年来，随着传感技术、电子技术和工业自动化的发

展，大部分的工业制造过程都可以被监测，从而产生了海量的数据，制造流程的复杂性和商业活动的频繁性同样产生了大量数据。如何利用生产过程中产生的数据，提取有价值的信息，从而改进生产过程，是智能制造主要需求，如针对重要设备的故障诊断、预测性维护和参数优化等。所有的智能技术都离不开对数据的分析，而数据分析任务是一种计算敏感型任务，需要大量的计算资源，这在工业环境中很难满足，尤其是以深度学习为代表的驱动技术。云计算虽然可以在远端提供一定的计算服务，但是由于时延不定、通信成本高等问题无法完全满足工业需求，因此边缘计算概念一经提出，就在工业环境中展开了广泛应用。本文总结了近几年在工业环境中的边缘计算应用，并从场景、动机和计算形式等方面进行了分析，相关信息与表 1 所示。

表 1 工业边缘计算应用文献综述
Tab.1 Literature review of industrial edge computing applications

文献	主要场景	动机					边缘计算形式
		安全性	低成本	低时延	低功耗	低流量	分布性
文 [21]	设备安全实时控制	✓					雾计算
文 [22]	工业物联网节点查找				✓	✓	雾计算
文 [23 - 25]	船厂管道加工可追溯性			✓			雾计算、微云计算
文 [26]	多智能体管理与控制			✓			雾计算
文 [27]	基于 LSTM 的刀具磨损监控			✓			雾计算
文 [28 - 29]	基于深度学习的产线工件识别与缺陷检测			✓		✓	雾计算
文 [30]	恶劣环境中基于移动机器人和深度学习的工业热异常检测			✓			雾计算
文 [31]	制造过程的实时数据分析和监控			✓			雾计算
文 [32]	智能制造过程中的设备运营与维护	✓		✓			雾计算
文 [33]	混沌生产场景中多智能体自主决策		✓	✓			雾计算
文 [34]	工业物联网中的深度学习应用			✓	✓	✓	雾计算
文 [35]	生产数据加密与溯源		✓				雾计算
文 [36]	智能电网数据加密	✓					雾计算
文 [37]	工业场景中机器学习应用	✓		✓			雾计算
文 [38]	生产物流中制造资源追踪定位		✓			✓	雾计算
文 [39]	基于 LSTM(long short-term memory) 的工业装备类别识别			✓			雾计算
文 [40]	基于无人机的工厂园区监控		✓	✓			雾计算
文 [41]	工业机器人网络化控制		✓	✓			雾计算
文 [42]	工业控制任务			✓			移动边缘计算
文 [43]	恶劣工业环境中传感器数据清洗		✓	✓		✓	雾计算
文 [44]	工业设备故障相关性寻找						雾计算
文 [45]	一种可以实现遥感、实时监控和可扩展的高性能数字孪生应用的信息物理机床		✓	✓			雾计算
文 [46]	矿山视频信息处理		✓	✓			雾计算
文 [47]	数字孪生中虚实同步		✓				雾计算
文 [48]	车载 CAN(controller area network) 总线入侵检测			✓			移动边缘计算

3.1 工业边缘计算主要应用场景

3.1.1 故障诊断与缺陷检测

边缘计算因为提供了更为便捷的计算资源，为工业现场数据分析任务提供了计算支撑，在所有的数据分析任务

中，故障诊断与缺陷检测类任务往往是最为重要的，因此也是目前应用最多的工业边缘计算场景。典型的应用有基于深度学习的刀具磨损监控^[27]和轴承的故障诊断^[49]、工厂产线零件识别与缺陷检测^[28-29]、工厂热异常检测^[30]、

智能制造过程的设备实时监控运维^[31-32, 50]、电力设备检修^[51], 这些应用普遍地利用了边缘计算低时延的特点, 提高了诊断预警的响应速度。而对于地理范围较大的故障检测, 边缘计算可以为移动设备与无人机等提供算力, 来提高检测效率^[52]。

3.1.2 工厂园区的安防监控

视频流的快速处理也是一种需要大量计算资源的任务, 厂区的视频监控对提高设备、物料、人员的安全性有着重要意义。因此, 基于边缘计算的视频流处理也是工业领域中重要的应用。如在矿山生产作业场景, 对视频数据进行结构化分析, 从而完成人员行为督导、设备状态监测、物料流转监控等任务^[46]。利用边缘计算实现铁路的无人坚守^[53]。边缘计算可以拓展园区安防的形式, 如通过边缘计算来为无人机提供计算服务, 进行大范围的数据收集和实时监控^[40]。

3.1.3 辅助设计与制造

虚拟现实和增强现实 (VR/AR) 技术近年来得到了快速发展, 在制造业中也有了一定的应用。VR/AR 技术通过数据可视化、交互性、沉浸性等特点可以辅助操作人员装配, 便于设备维护和产品测评等^[54]。由于 VR/AR 技术需要处理大量信息, 并且一些穿戴式 VR/AR 设备难以提供强大的算力, 因此边缘计算在此类应用上显得尤为重要。西班牙 Navantia 造船厂利用边缘计算技术来为船厂的 AR 设备提供计算服务, 从而提高了 AR 设备的响应时间, 更好的帮助工人快速的处理信息^[25, 55]。

数字孪生^[56]是辅助设计与制造又一重要方法, 它将物理实体进行数字化, 通过数字模型来模拟物理实体在真实环境中的行为, 来进行虚拟与现实的不断交互反馈, 从而更新产品的设计水平。文[47]利用边缘网关来实现了数字孪生中的虚实同步技术, 设计了一种基于数字孪生的生产单元可视化管控方法。文[45]为提高数字孪生应用的计算效率, 缩短映射延迟、减少云端计算量, 提出了一种基于边缘计算的信息物理机床, 实现了实时遥感、实时监控和可扩展等高性能数字孪生应用。

3.1.4 工业数据挖掘

边缘计算提供了大量的分布式计算节点, 这些节点除了提供计算服务外, 还可以提供相关的地理位置信息, 这些位置信息对于工厂流水线生产过程中大量工件的定位与溯源有着重要作用。如西班牙 Navantia 造船厂利用雾节点、微云节点与蓝牙、RFID 技术结合, 实现了船厂内数万计管道的定位识别、质量评估与溯源等功能, 提高了管道车检的产量^[23-24]。在工业生产线上, 流动的除了产品还有一些数据信息, 如设备的故障信息。文[44]利用边缘计算来捕获异常和故障在传感器及设备直接的传播, 获取故障的相关性信息, 从而进行工业设备的预测性维护。文[43]利用边缘计算传感器数据清洗, 在边缘节点通过 SVM 建立数据清洗模型, 在保证数据的可靠性和完整性的同时, 降低了传送到云端的带宽和能耗。

3.1.5 控制决策过程的优化

智能制造^[1]的提出使得工业生产过程变得日益复杂,

这对工业生产的控制与决策过程提出了更高层次的优化需求, 以深度学习为代表的复杂优化方法在工业控制领域也有着较多的应用。边缘计算可以为这些应用提供基础的计算设施, 保证相关的计算任务安全、快速、高效地完成。文[33]针对随机顺序的混沌生产场景, 利用边缘智能构建了多个智能体系统, 通过不同的智能体之间的行为交互提高自主决策能力, 从而提高生产过程的自适应性和鲁棒性。文[57]提出了基于边缘计算的云机器人系统, 利用边缘计算提供的实时计算能力, 提高了机器人的自主移动能力、感知推理能力。文[41]提出了一种基于雾计算的工业机器人局域网络协同控制体系结构, 并在雾计算网络中设计了一种基于 RNN (recurrent neural network) 的多 AGV (automated guided vehicle) 协调策略, 比传统的云机器人提高了 35% 的效率。文[58]提出了一种基于 5G 和边缘计算的智能汽车柔性制造解决方案, 提高了柔性制造中感知、分析、决策和执行过程的效率。

3.1.6 工业数据安全性与隐私保护

工业互联网的发展使得越来越多的设备数据都需要与工业云平台进行交互, 用户数据往往携带大量的用户隐私, 与云端交互的过程存在用户隐私泄露的问题。工业设备与云端通信的过程也可能被恶意第三方入侵, 相比于传统互联网的网络入侵, 工业互联网由于数据源可能是大型设备, 入侵更容易造成严重的安全事故, 因此安全性与隐私保护一直是工业安全研究的重点。边缘计算在云端与设备端之间提供了多级计算资源, 为工业应用的安全和隐私保护提供了更灵活的方法^[59]。针对隐私泄露问题, 可以利用边缘节点对采集的数据进行加密压缩^[21, 60-61]、多点聚合^[62]来保护原始数据的特征不被轻易获取, 或者直接将云端计算下放到边缘端来执行, 减少不必要数据的上传^[62], 从根本上杜绝隐私泄露的可能。特别地, 针对深度学习应用, 联邦学习^[63]可以通过模型聚合的方法来取代原始数据的上传, 从而避免用户隐私泄露, 这种方法非常适合在边缘计算体系下进行^[64]。对于恶意入侵导致的安全问题, 可以对众多的边缘设备进行注册认证, 只有合法注册的设备才能收发数据^[62, 65], 同时边缘计算也适合与区块链技术进行融合, 来解决工业互联网安全性和隐私保护的问题^[66]。

3.2 边缘计算在工业环境中应用的主要动机

边缘计算在工业场景中应用的主要动机如表 1 所示。与云计算相比, 边缘计算的低时延是工业场景中主要考虑的动机, 如设备的故障诊断快速预警、工业控制任务都需要较低的响应时延。对于大量数据的分析, 尤其是视频图像信号、高频振动信号等, 云端处理会浪费大量的通信流量消耗, 通过边缘计算可以降低大量的云端通信量, 从而降低通信成本, 而将部分计算卸载到边缘侧进行, 也降低了云计算的使用, 从而节省了计算成本。安全性是边缘计算在工业领域应用的又一动力, 工业生产数据通常携带大量用户隐私, 原始数据的上传会带来严重的安全问题, 边缘计算可以通过本地加密^[35-36]在一定程度上解决该问题。同时, 边缘计算又可以通过多级签名加密认证, 从而减少

恶意入侵的几率,提高设备安全性,这在工业应用中是尤为重要的.边缘计算是一种分布式计算,这种分布式的特性同样推动其在工业中的应用,例如分布式可以提供更好的感知定位功能,方便产品的追踪溯源等^[23-25,44].

3.3 西门子工业边缘平台

西门子自从2018年推出第一台工业边缘设备以来,其边缘计算技术经过不断的完善发展,已经形成了一个较为完备的工业边缘计算平台.包括用于执行边缘任务的各种工业级边缘设备(SIMATIC IPCs, SIMATIC S7-1500, RUGGEDCOM RX1500等),基于docker容器构建的工业边缘运行时,一些用于监控、性能分析、数据分析、性能分析软件、仿真

设计的工业边缘软件,用于管理各种边缘设备的管理系统及工业云平台 MindSphere 等.西门子提供了一个通过其工业边缘计算平台构建的数控机床案例,如图2所示.在云平台中,设备制造商用工业云平台对全球设备数据进行评估和分析;工业边缘管理系统运行在常规的PC上,对应用软件和众多边缘设备集中管理,协调软件更新与扩展;可以通过 OPC UA 灵活连接 MES(manufacturing execution system) 和 ERP(enterprise resource planning) 系统;工业边缘 APP 运行在边缘设备上,为生产人员提供设备效能,边缘设备通过 PLC(programmable logic controller) 连接控制数控机床.

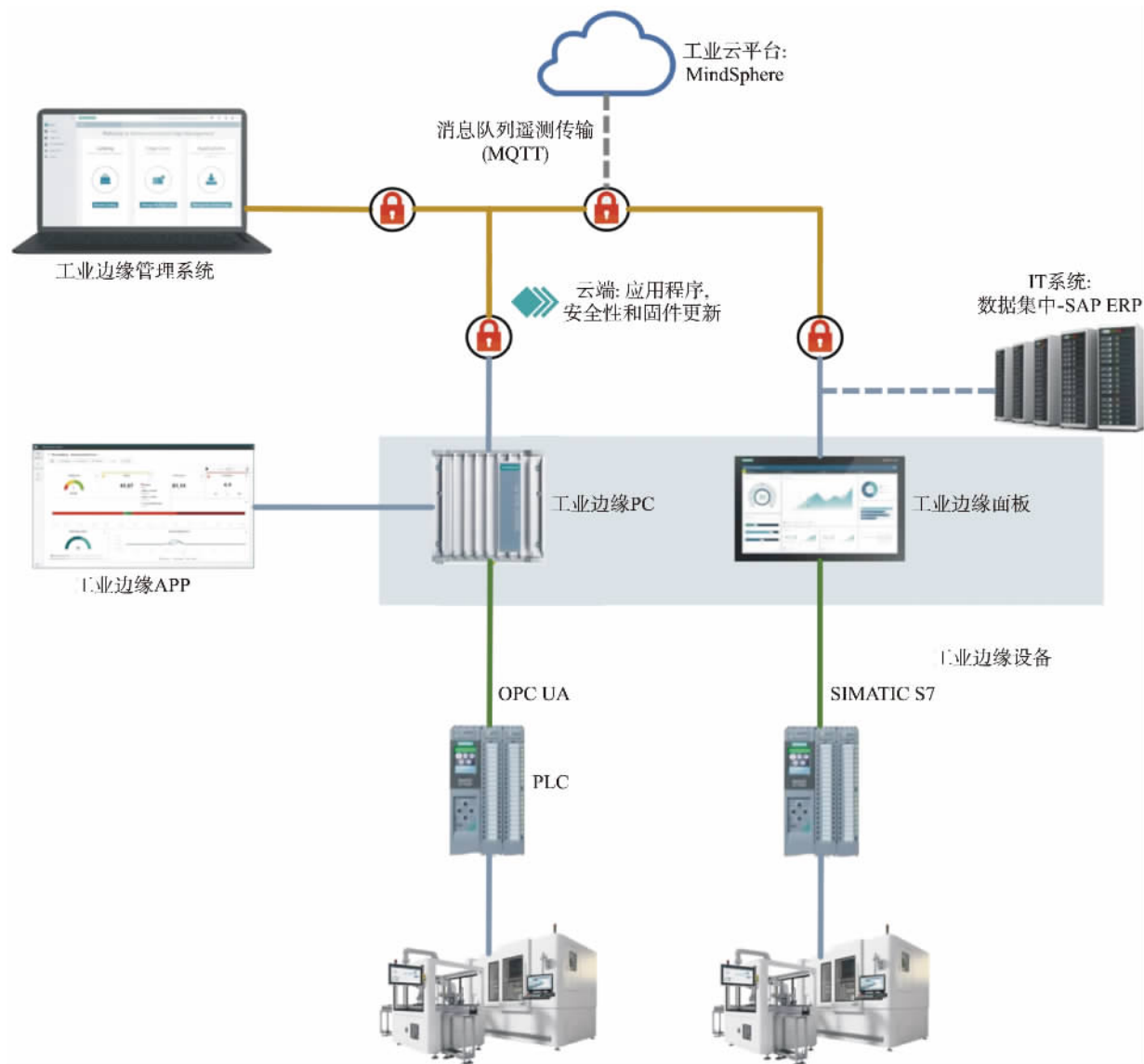


图2 西门子工业边缘平台示意图

Fig.2 Schematic diagram of Siemens industrial edge platform

3.4 华为智能边缘平台 IEF

华为的智能边缘平台 IEF 现在主要的工业应用场景为面向工业产线的视觉缺陷检测,用边缘智能的方式取代产线人工的质检,从而提高检测的效率.平台的结构如图3

所示,边缘节点部署在工业产线,采集工业现场的图片信息,利用部署在边缘节点的视觉 AI 模型进行识别,同时在边缘节点生成相关的数据库. IEF 与华为云相结合,在云端完成模型的训练并下放的边缘节点进行部署,并可以进

行增量的训练优化。不同于西门子的工业边缘平台开发了大量的工业级边缘设备, 华为的边缘节点硬件为满足一定

配置的 PC, 华为 IEF 主要提供边缘计算软件与云端的服务, 边缘节点的软件同样采用 docker 容器的方式进行部署。

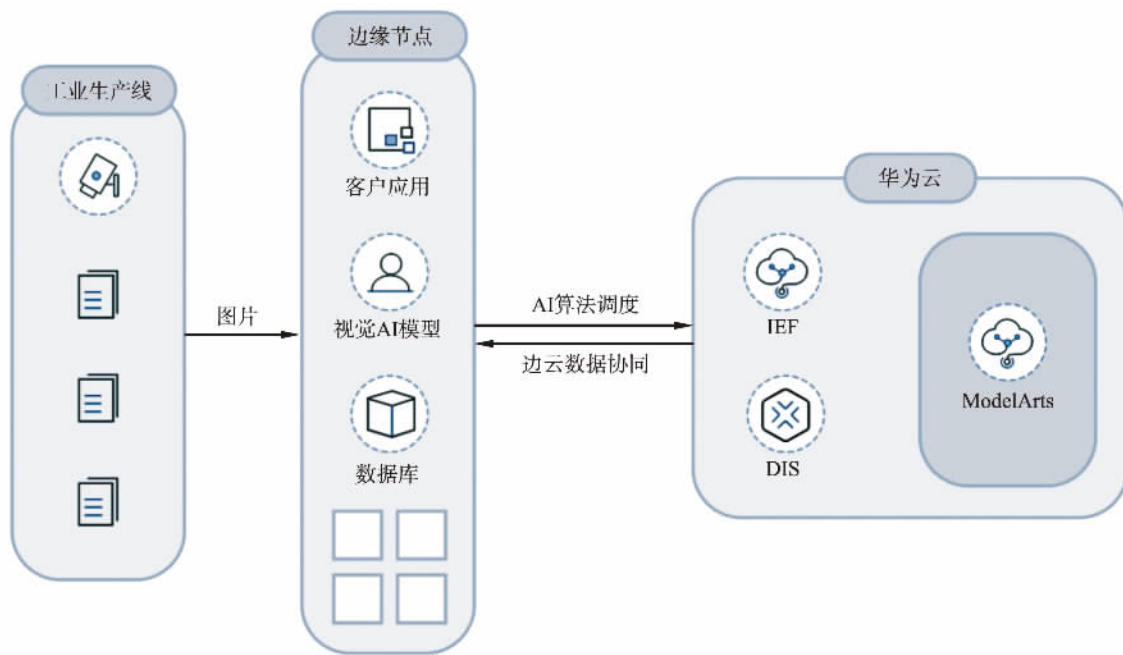


图3 华为的智能边缘平台 IEF 示意图

Fig.3 Schematic diagram of Huawei's intelligent edge platform IEF

3.5 工业互联网平台边缘计算标准化进展

由于制造业的复杂化和多样化, 在实际工业应用中落地中对边缘计算需求不同, 导致各厂商边缘计算平台软硬件差异巨大, 设备接口、数据标注难以兼容。针对该问题, 2020年5月, 工业互联网产业联盟发起了“边缘计算标准计划”, 联合各方展开设备与平台标准研制、评估测量等工作, 该计划共针对7种工业互联网边缘计算标准立项:

- 《工业互联网边缘计算总体架构与要求》
- 《工业互联网边缘计算节点模型与功能要求: 边缘控制器》
- 《工业互联网边缘计算节点模型与功能要求: 边缘网关》
- 《工业互联网边缘计算节点模型与功能要求: 边缘云》
- 《工业互联网边缘计算技术要求与测试方法: 边缘控制器》
- 《工业互联网边缘计算技术要求与测试方法: 边缘网关》
- 《工业互联网边缘计算技术要求与测试方法: 边缘云》

4 边缘计算的关键技术

边缘计算的主要研究目标就是为了给计算任务提供更为便捷的计算服务, 包括低时延、低流量、低成本、分布广、安全、可靠等优点。而在工业应用中, 边缘计算的研究应该考虑到工业场景的特点, 包括任务类型、计算需求、软硬件体系、通信方式等。为此, 本文对近年工业场景中的边缘计算关键技术的研究现状进行了综述, 并分析

了工业应用的技术难点与未来的研究方向。

4.1 计算卸载

计算卸载的主要目的是将本地的计算任务卸载到云端或者边缘端等不同的位置进行执行, 以完成不同的目标的优化, 卸载目标可以分为降低任务时延、降低设备能耗、最小化计算成本或者混合目标。计算卸载需要对计算任务进行建模, 划分成多个子任务, 然后将根据优化目标决策子任务的执行位置。计算卸载与后续讨论的资源分配存在耦合的部分, 因为计算卸载在部分情况下也会考虑子任务的资源分配, 但是计算卸载的主要研究对象是计算任务, 而资源分配主要研究对象为计算设备, 因此分开讨论计算卸载与资源分配。

4.1.1 卸载模型

对于计算卸载, 任务建模尤为重要。完整的任务模型可以更好地分析任务之间的计算依赖关系, 并且允许更细粒度的卸载, 从而更容易寻找到计算任务卸载的最优解。带权重的有向无环图(directed acyclic graph, DAG)^[67-70]是现阶段描述计算任务最常见的模型。如图4所示, 一个计算任务被描述成DAG, DAG由带权重的边和节点组成, 节点表示计算过程, 边表示计算过程的传递依赖关系。其中, 根据卸载场景的不同, 节点的权重 c_i 可以被建模为计算过程的资源消耗, 如CPU、GPU、内存、能耗等; 边的权重 t_i 可以建模为通信过程中的资源消耗, 如带宽、流量、能耗等。越细粒度的建模, 所生成DAG的节点越多, 并且相应的权重越低。对于原始模型转化为DAG进行卸载, 往往很难取得较好的卸载效果, 因为计算量和数据量往往是

正相关的, 计算卸载在降低计算时延的同时可能会带来更多的通信时延. Ren 等^[71]提出了一种计算压缩的卸载模型, 可以降低卸载过程中带来的通信消耗, 从而降低卸载时延.

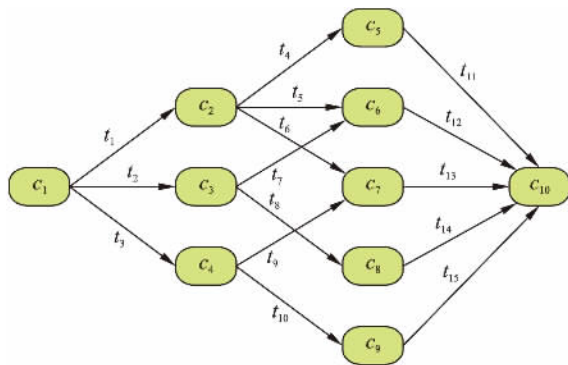


图4 计算任务 DAG 建模图

Fig.4 DAG modeling diagram of computing task

4.1.2 不同目标的计算卸载研究

边缘计算最主要的目的是为了解决云计算时延的问题, 因此, 降低任务时延往往是首先考虑的卸载目标, 包括平均时延^[72-73]、考虑长期的任务执行的累积时延^[74]与最坏时延等^[75]. 通常采用的方法有根据边缘侧计算资源与卸载到云端的通信资源来计算任务的执行与发送时间, 来决定卸载策略从而获得最小推理时延.

实际场景中, 由于进行卸载的终端多是电池供电的移动终端或者是功率受限的设备, 所以终端的能耗也是卸载的一个重要目标. Jiang 等^[76]在考虑计算端到端的时延约束前提下, 提出了一种节能的卸载决策机制. Liu 等^[77]在权衡本地计算和已经分配的卸载资源同时, 提出了一种动态任务卸载方法来最小化用户能耗. Wu 等^[78]针对移动设备有限的电池容量和长期运行需求之间的矛盾, 设计了一种节能的计算卸载策略.

因为边缘计算与云计算的计算成本不同, 云端计算费用往往要高于边缘计算, 因此, 计算成本通常也是卸载考虑的目标之一. Baranwal 等^[79]在综合考虑了云计算、微云计算和边缘计算的前提下, 利用博弈论建立了一种非合作环境下的任务卸载决策模型, 并通过 iFogSim 进行仿真实验, 证明了该方法可以最小化任务计算成本.

为了满足计算任务的复杂要求, 还有相当多的研究将上述的卸载目标进行联合考虑. 如综合考虑任务时延、设备能耗和计算成本的同时, 对任务进行卸载^[80-83].

4.1.3 深度学习任务的计算卸载

边缘计算的目标是为计算任务提供多元化的计算资源, 因此计算敏感型任务将是边缘计算中最主要任务负载, 其中典型的代表就是深度学习算法. 深度学习任务的卸载, 主要目的是选择一个合适的分区点来分离模型, 决策模型的执行位置. 为了使深度学习任务更适合进行边缘卸载, 往往需要对模型进行改进, 如 BranchyNet^[84]、多分支神经网络 (MBNN)^[85], 多数的云边分区的卸载方案都是

利用的这种结构^[86-90]. 文[86]将云边协同的深度学习卸载方法应用到了工业生产线的故障检测中, 使中央服务器的计算负载转移到边缘节点, 提高了计算效率. 文[91]通过知识蒸馏来进行边缘侧的深度学习模型轻量化, 来方便模型在边缘端部署.

4.2 资源管理

资源管理指的是对边缘计算环境中的资源进行管理和优化的技术, 资源具体的可以是 CPU、GPU、内存、IO、硬盘等计算资源或者带宽、频谱等通信资源. 资源管理是决定边缘计算水平的重要技术, 传统的云计算对资源管理技术已经有了很多研究, 但是云计算中计算设备类型单一, 并且资源水平也基本相同, 而工业边缘计算环境中, 计算设备在所有者、计算架构、资源水平、计算成本、通信成本等方面存在较大的异构性, 因此, 工业边缘计算中的资源管理往往比云计算中的资源管理存在更多的问题与挑战.

边缘计算的资源管理按照具体的功能可以分为资源放置、资源分配和资源定价与激励. 资源放置指的是在一定地理范围内部署边缘节点, 包括设置边缘节点的位置、计算资源水平、通信资源水平等. 资源定价与激励指的是为资源或服务设定合理的价格, 而激励问题考虑到资源使用者同时又是贡献者, 如何激励用户将自己的资源贡献出来参与计算, 是资源激励主要的研究问题. 资源分配问题是资源管理中最常见的问题, 资源分配指的是为计算任务分配相应的计算与通信资源, 这是边缘计算资源管理中研究最多的内容.

4.2.1 资源放置

资源放置在边缘计算的设备部署阶段, 场景可以是在某个地理范围内部署微云服务器, 也可以是在某个厂区内部署边缘服务器或者边缘网关. 资源放置根据场景不同有着不同的研究目标, 如最小化部署成本、最小化计算时延等. Xu 等^[92]和 Ren 等^[93]研究了在由多个无线接入点构成的大规模无线城域网环境中的相同的微云放置问题, 完成最小化所有用户和服务器之间的平均时延^[92]和满足任务需求的同时最小化服务器放置数量来最小化部署成本的部署目标^[93]. Lin 等^[94]针对工业 4.0 的智能物流场景中, 对工业边缘网关和服务器进行了部署, 在满足最大计算容量、最大计算时延及覆盖范围的约束前提下, 使得部署成本最小. 该工作首次考虑了在工业传感器布局基础上的异构计算设备部署问题. 文[95]对无人机搭载的边缘服务器的位置部署进行了研究, 使其可以覆盖更大范围的移动终端, 但是由于无人机总功耗的限制, 此类研究在现阶段电池技术没有取得重大突破的前提下很难具有实际意义.

4.2.2 资源定价与激励

资源定价和激励是将计算资源以货币或者其它形式进行量化, 好的定价和激励的策略有助于获取更多的计算效益. 对于微云计算和移动边缘计算这种有运营商统一部署的边缘计算形式, 定价问题是其主要研究内容, 对于像雾计算这种用户既参与计算, 又贡献资源的计算形式, 资源激励是其研究内容.

拍卖模型是资源定价的主要模型^[96-97]等研究都采用拍

卖模型进行资源的定价, Jin 等^[97]提出了一种用于微云定价的激励兼容的拍卖机制(ICAM)。ICAM 可以有效地分配微云的计算资源以满足移动设备的需求, 并保证了个体合理性、预算平衡、买卖双方的真实性及计算效率。Sun 等^[96]针对工业物联网中具有限制性资源的边缘计算服务器的资源拍卖问题, 提出了一种双向拍卖的方案, 首次考虑到边缘节点的资源限制性, 并通过计算资源动态定价, 提高了系统的计算效率和收益。Sun 等^[98]针对零散的分布式雾节点, 提出了一种整合网络中闲置资源的众筹算法, 同时为了鼓励更多资源参与雾计算, 提出了一种激励机制, 可以降低服务级别协议(service level agreement, SLA)的违反率。

4.2.3 资源分配

资源分配的研究内容是如何为边缘设备上的计算任务分配相应的资源来完成一定的优化目标。资源分配的研究往往根据场景选择其中的一种或者几种计算资源进行分配, 如对于计算敏感型任务往往会考虑 CPU、GPU、内存等计算资源分配, 数据敏感型任务会考虑写入/读取(I/O)、硬盘、频谱和带宽等存储通信资源进行分配。

资源分配研究的目标有很多, 通常有最小化计算时延^[71, 99-104], 最小化系统能耗^[105-108], 最小化计算成本^[109-113], 提高资源利用率和系统效率等^[114-118], 资源分配的研究目标根据场景不同可以是上述研究目标的一种或者几种。在任务执行过程中, 分配的计算资源不进行变化的分配叫做静态分配, 任务分配的计算资源动态变化的分配叫做动态分配。通过任务资源的动态分配, 可以实现任务的调度。资源分配需要以计算资源虚拟化技术为载体, 最常见的虚拟化技术有容器和虚拟机, 也有少数研究以 Java 虚拟机^[119]和线程^[120]来进行资源分配。虚拟化是边缘计算重要的技术之一, 轻量级的虚拟化技术有助于降低计算时延、减小系统消耗, 关于虚拟化技术在第 4.3 节中单独讨论。

4.3 计算虚拟化

计算虚拟化技术是一种计算机领域常见的资源共享技术, 它可以将软硬件资源按照不同的级别进行虚拟化, 使得不同的应用程序在虚拟的设备上运行。计算虚拟化技术可以共享设备资源, 提高设备利用率, 有助于应用程序的灵活开发部署, 是云计算和边缘计算的实现基础, 计算卸载、资源管理都需要以计算虚拟化技术为载体。按照虚拟化等级由低到高分类, 计算虚拟化可以分为指令级别虚拟化(QEMU)、硬件级别虚拟化(VMware 虚拟机)、操作系统级别虚拟化(Docker 容器)和应用程序级别虚拟化(JVM)。虚拟化等级越高, 其轻量化程度越高, 但应用程序部署的灵活性越低。在云计算和边缘计算中, 虚拟机技术和容器技术是最常见的两种虚拟化技术。

4.3.1 虚拟机技术

虚拟机(VM)是硬件级别的虚拟化技术, 它可以将宿主计算设备的实际硬件虚拟化出多个不同的硬件, 并在上面运行不同的操作系统, 来部署应用程序。Cloudlet^[14]是最早通过虚拟机技术实现的边缘计算的形式。Cloudlet 中虚拟机部署的时延约需要 1 min ~ 2 min, 这很难满足实际

的应用需求, Ha 等^[121]通过重复数据的删除来减少 VM 覆盖层的大小, 从而将 VM 动态合成加快到 10 s 左右。Lu 等^[122]的针对 VM 的远程动态加载加速, 提出了一种集成跟踪预测方法来集成虚拟 I/O 以优化启动时间。该方法通过跟踪历史访问模式来生成服务启动过程的应用加载请求列表, 并将列表中关键 I/O 请求预取并自动缓存, 以加快服务启动时间。由于软件技术的发展和制造过程的日益复杂度, 虚拟化技术由于其便于开发部署在工业环境也得到了广泛的应用, 如文^[123]实现了一种支持 IEC 61131-3 标准的 PLC 虚拟机, 可以让 PLC 开发人员快速的将符合标准的程序应用到不同的平台, 方便加载各种外围驱动程序和用户配置。Breivold 等^[124]针对利用虚拟化硬件进行系统测试进行了评估, 介绍了通过虚拟化的方式进行自动化测试可以更好地创造伸缩的测试环境、注入硬件故障等。

4.3.2 容器技术

容器技术是一种更轻量级的计算虚拟化技术, 它在操作系统级别进行虚拟化, 所有容器共享操作系统内核, 因此具有更低的资源需求和启停时间。因为边缘计算的资源限制性, 使得容器技术在边缘计算中得到了广泛的应用。卡耐基梅隆大学的 Elgazar 等^[125]提出了一个 docker 容器动态迁移的框架 Teddybear, 其架构如图 5 所示, 整个系统由云端实体、多个可以运行 docker 的边缘计算及用户移动设备组成。移动设备可以通过云实体或者 P2P 的方式将本地的 docker 容器迁移到边缘端执行。而对于资源受限的嵌入式平台, 依然可以通到 docker 来实现边缘计算功能。Avino 等^[126]对于基于 docker 的边缘计算消耗进行了量化, 对在不同场景中适用性分析提供了帮助。Ismail 等^[127]从部署与停止、资源与服务管理、容错与缓存等方面对基于 docker 的边缘计算进行了评估, 证明了 docker 可以提供快速部署、轻量化内存占用和良好的性能。Bellavista 等^[128]评估了 docker 在树莓派上的可行性, 证明了通过适当的配置和优化, docker 可以实现嵌入式平台上进行灵活简单的部署。

容器技术在工业环境中也有着广泛的应用, 如埃克斯美孚倡导的开放式流程自动化(OPA)^[129]。博世公司^[130]针对制造企业缺乏便捷的对生成数据分析、优化生产过程的平台, 开发了一种基于 Docker 微服务的数据分析平台, 为制造业是数据分析提供便利, 简化了不同数据源的数据合并使用, 提供便捷的数据分析服务, 并方便在边云环境中部署。针对灵活复杂的生产过程对 PLC 控制的灵活性需求, ABB 公司^[131]提出了一种基于 Docker 容器技术的多用途控制器体系结构, 包括对多个 PLC 执行引擎的支持, 从而使 PLC 控制器可以在工业生产中进行灵活的部署。

4.3.3 开源边缘计算容器管理平台

Docker 容器在实际生产环境中使用需要依赖相关的容器管理平台来对容器进行自动化部署、伸缩、管理监控等, 主流的 docker 管理平台有如 Kubernetes^[132]、Docker swarm^[133]等。这些平台都在云计算中得到了广泛的运营, 但是在边缘计算环境中, 由于资源的限制性, 上述平台难以应用, 以 Kubernetes 为例, 其节点最低配置需要 4 核和 16 GB 的内存。针对上述需求, 许多公司和组织都针对该需求在

Kubernetes 的基础上开发了轻量化的 docker 管理平台, 主要代表有 Rancher 公司的 K3s^[134] 和华为公司的 KubeEdge^[135]. K3s 是一个轻量化的 Kubernetes, 为了更好地跟云平台相兼容, K3s 完全兼容 Kubernetes, 并打包为单个二进制文件, 只需要 512 MB 的内存即可运行, 并可以在 ARM 架构的嵌入式平台上运行. K3s 的架构如图 6 所示, 分为 Server

节点和 Agent 节点. 与 K3s 将 Kubernetes 轻量化的方向不同, KubeEdge 在云端仍然采用 Kubernetes, 其架构如图 7 所示. KubeEdge 通过部署边缘控制器及云边 Hub 来将 Kubernetes 扩展到边缘侧, KubeEdge 的部署需要云端的 Kubernetes 支持, 可以更好进行云边协同, 而 K3s 可以仅在边缘端进行自治, 部署更加灵活.

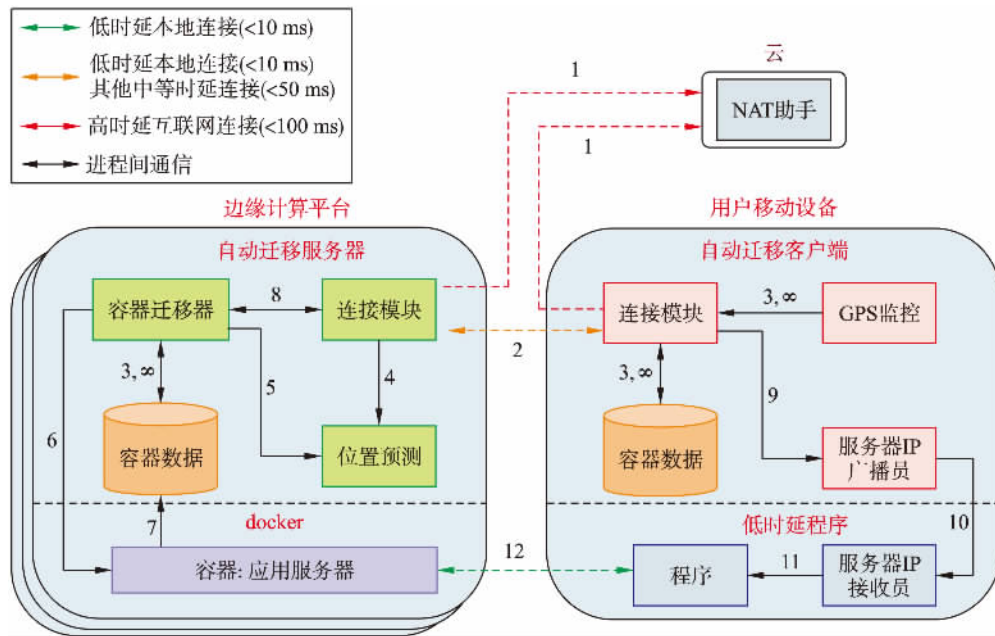


图 5 Teddybear 架构图

Fig.5 Teddybear architecture diagram

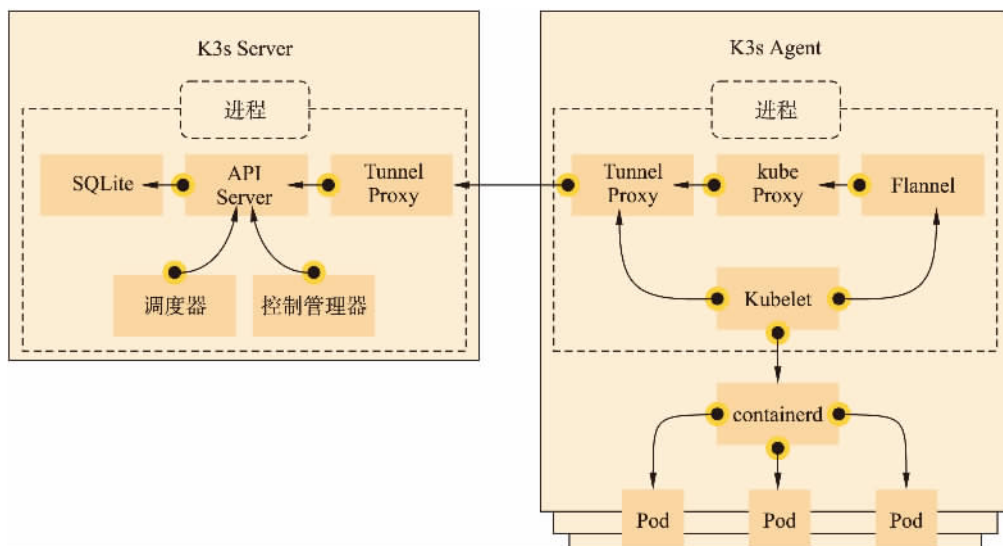


图 6 K3s 架构图

Fig.6 K3s architecture diagram

4.3.4 其他虚拟化技术

Messenger 等^[136] 将应用程序虚拟化作为计算卸载中更轻量化的方案, 定量地分析了该方案在时延和能耗方面的性能指标, 定性比较了应用程序虚拟化与 VM 动态合

成, 证明在部署阶段, 应用程序虚拟化更有优势. Wu 等^[137] 提出了一种基于 Android Unikernel 的计算卸载方案, Android Unikernel 相比于容器和虚拟机更轻量化, Wu 等提出了一种 Rich-Unikernel 的概念, 旨在一个单内核中支持

各种应用程序, 同时避免了重新编译.

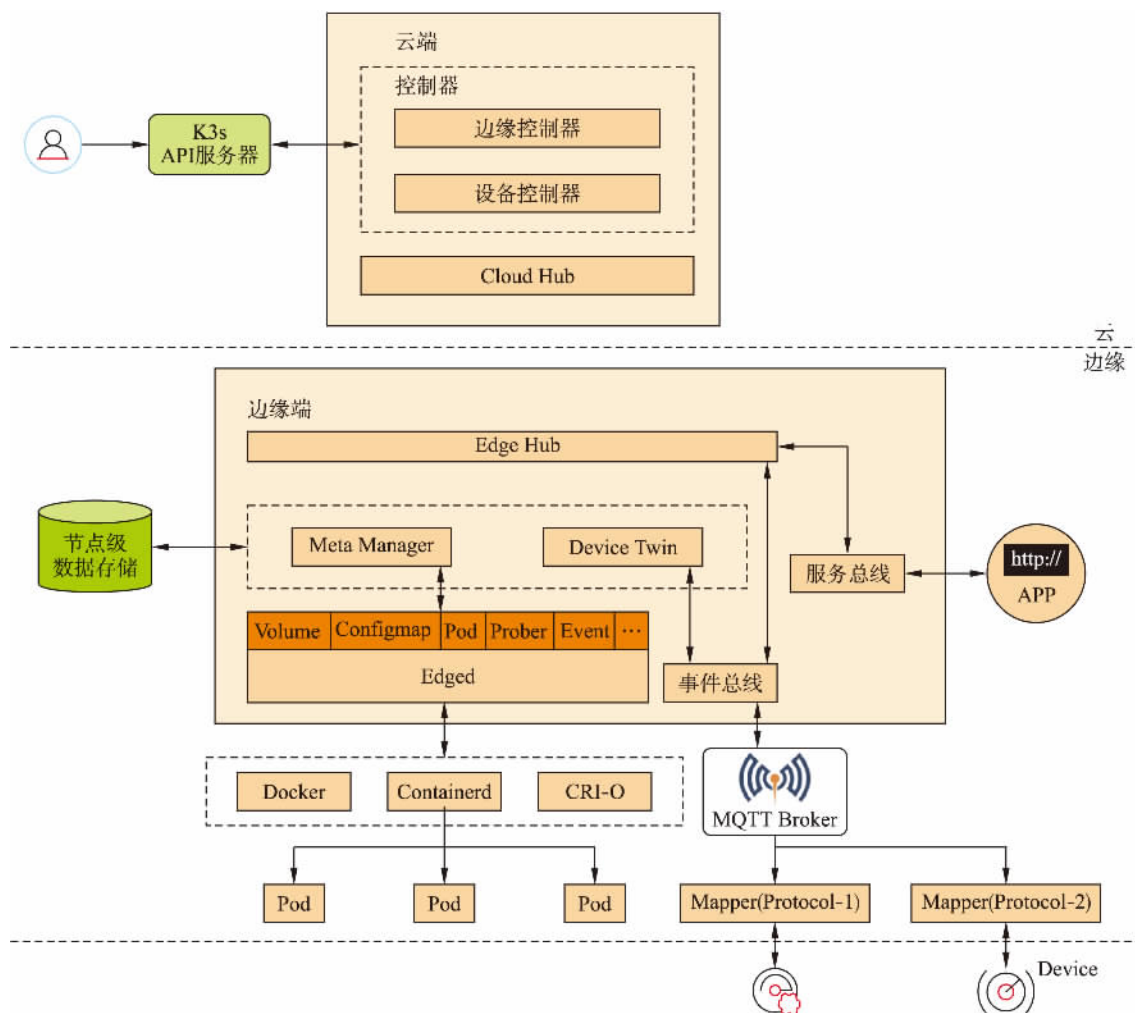


图7 KubeEdge 架构图

Fig.7 KubeEdge architecture diagram

5 工业边缘计算的研究挑战

5.1 确定性时延的边缘计算系统建模

现阶段中对边缘计算的研究, 往往将低时延作为优化目标或者约束条件, 但是在工业环境中, 由于工业任务往往具有严格的程序控制的, 不同于人类参与的环境, 较低的时延可以获取较高的 QoS 体验. 工业环境中一味的追求低时延不会为工业过程带来更好的效果, 例如在控制任务中, 控制周期一定, 低于控制周期完成计算任务往往没有实际意义并且会带来计算浪费. 因此在工业计算任务, 确定性的计算时延往往比更低的计算时延更有意义, 但是目前尚未有针对确定性时延的相关研究. 确定性时延的系统建模应考虑系统中所有元素完备的确定性时延, 即从任务生成端到边缘设备执行端再到执行结果返回端数据流经的所有通信计算都应该是确定性时延的, 具体应包括实时操作系统、实时虚拟化技术、确定性时延的通信, 如 5G 与 TSN(时间敏感网, time-sensitive-network), 然后考虑针对不同时延要求的任务数据集进行计算卸载与资源管理, 将

任务时延抖动建模为系统约束, 根据实际需求对成本、能耗、效率等一个或者多个目标进行优化.

5.2 考虑边缘计算形式工业任务设计与卸载研究

现阶段关于工业场景中的计算卸载其计算任务大多不是针对边缘计算形式设计的, 多是传统的嵌入式任务或云计算任务, 此类任务在设计时未考虑分布式的计算场景, 因此卸载此类任务往往难以取得较好的效果. 以深度学习应用为例, 传统的模型以云边端协同方法卸载到云端, 因为模型中间层传递的数据量往往是大于输入和输出层的, 这导致最优的卸载点往往是本地或者云端. 而相关改进的模型^[83-84]适合在云边环境中进行卸载, 尤其是文[84]针对工业故障诊断场景设计了一种快速的神经网络结构, 特别适合在资源水平不一的工业云计算、边缘计算的环境中批量部署. 未来针对工业场景的计算卸载, 应该考虑改进或者重新构建适合工业云边计算架构下的计算任务, 然后进行卸载研究.

5.3 异构计算环境的资源管理研究

工业场景下的边缘计算明显不同的地方就是边缘设备

和计算任务的异构性较强. 工业边缘设备组成包括工业 PC、服务器、各种嵌入式设备, 它们在资源水平、系统架构、能耗要求等方面存在着较大的差异. 而工业应用中的计算任务有控制、通信、显示、报警、训练等多类组成, 它们有着不同的计算需求, 如时间敏感型的控制任务、报警任务, 计算敏感型的控制任务等. 如何在这种异构计算资源环境下, 对异构计算任务的计算资源进行动态匹配, 在满足计算任务约束的前提下, 使得系统性能最优, 这是工业场景下边缘计算研究的一个挑战.

5.4 限制性资源环境中的 GPU 管理

计算敏感型任务是边缘计算中主要的一类任务负载, 深度学习类任务是其中主要的代表, 在工业环境中也有着广泛的应用^[138-139], 这也成为边缘计算环境和中的主流负载, 但是目前的研究中, 计算敏感型任务通常被建模为 CPU 和内存敏感型, 而未考虑到 GPU 设备, 这种建模方法并不严谨. 随着深度学习技术的不断应用, GPU 在并行计算中占据了主导地位, 在深度学习的相关运算中, 特别是资源限制性的嵌入式环境中 GPU 的效率远高于 CPU, 以 NVIDIA Jetson Nano 平台为例, 相同的计算, GPU 速度比 CPU 快近百倍^[84]. GPU 虚拟化技术是 GPU 边缘计算的支撑技术之一, GPU 虚拟化技术在虚拟机中有着广泛的研究^[140-141], 在容器中, 用户可以通过 NVIDIA-docker^[142] 在容器环境下使用 GPU. 但是在容器环境中应用仍然处于起步阶段, 现阶段的应用通常只考虑在容器中使用 GPU, 很少考虑多个容器间的 GPU 共享与管理, 并且其管理单元往往为物理显卡, 无法实现更细粒度的显存管理. 这是未来应用的挑战之一, 尤其是在工业环境中, 更好的利用嵌入式 GPU 来处理计算密集型任务可以获得更高的计算效率.

5.5 实时容器技术

对于虚拟化在工业领域中的应用, 实时性仍是最大的挑战. ABB 公司^[143]从应用的角度对容器和虚拟机的实时性进行了评估, 分析了宿主机操作系统对容器和虚拟机启动时间的影响. 斯图加特大学的 Luca Abeni 等^[144]研究了控制任务容器化部署的实时性问题, 通过将 Linux 的实时补丁 PREEMPT_RT 与 Cobalt 内核融合, 测试了消息的往

返时延在 500 μs 级别, 比萨圣安娜大学 Melcher 等^[105]提出了一种利用 Linux 内核中实时截止日期的调度策略来保证容器的实时性问题, 将 SCHED_DEADLINE 调度策略扩展到 Linux 控制组并与 Docker 兼容, 实验证明了毫秒级实时调度任务的可行性. 实时容器技术是工业边缘计算中重要技术, 但是现阶段边缘计算往往未考虑实时容器技术的应用, 包括所有以时延为约束的优化也未考虑操作系统与虚拟化本身带来的时延. 基于实时容器技术的虚拟化技术研究有助于边缘计算在工业领域中落地.

5.6 更广泛的资源共享形式

基于容器或者虚拟机的虚拟化是边缘计算的主流的计算资源共享技术, 但是这种形式的虚拟化对边缘设备有一定的性能要求, 需要操作系统的支持. 对于工业场景中大量不支持操作系统特别是不含有内存管理单元 (memory manage unit, MMU) 的硬件设备, 如单片机等, 很难通过虚拟化来实现边缘计算, 因此需要一种更广泛的资源共享来解决该问题. 基于模型设计 (model-based design, MBD) 中的代码生成是一种可行的解决方案, 这种设计方法已经在机器人、汽车电子领域得到广泛的应用. 传统的 MBD 方法的代码生成只面向与单片机, 未考虑面向不同类型的设备进行代码生成, 如果在 MBD 设计阶段, 模型面向包括单片机、虚拟机、容器在内的不同边缘设备设计, 并通过代码生成的方式生成相关的代码, 则可以实现支持更多类型设备的边缘计算.

6 总结

本文首先介绍了工业互联网的技术背景, 分析了边缘计算的起源与定义, 介绍了几种典型的边缘计算形式与工业环境下边缘计算架构. 对近几年工业场景中边缘计算的应用进行了综述, 分析了边缘计算在工业应用中的主要动力与一些对边缘计算需求较大的工业场景, 并介绍了现有的工业边缘计算平台与标准化进程. 分析了对工业应用影响较大的计算卸载、资源管理及虚拟化等边缘计算关键技术, 最后对边缘计算工业应用研究的不足与挑战进行了分析, 希望从而可以给相关的科研与工程技术人员提供一定参考.

参考文献

- [1] 周济. 智能制造——“中国制造 2025”的主攻方向[J]. 中国机械工程, 2015(17): 2273-2284.
Zhou J. Intelligent manufacturing-main direction of “Made in China 2025”[J]. China Mechanical Engineering, 2015(17): 2273-2284.
- [2] Lee J. A view of cloud computing[J]. International Journal of Networked and Distributed Computing, 2013, 1(1): 2-8.
- [3] Malhotra L, Agarwal D, Jaiswal A. Virtualization in cloud computing[J]. Journal of Information Technology & Software Engineering, 2014, 4(2). DOI: 10.4172/2165-7866.1000136.
- [4] Kavitha B, Varalakshmi P. Performance analysis of virtual machines and docker containers[M]//Communications in Computer and Information Science, vol.804. Berlin, Germany: Springer, 2017: 99-113.
- [5] Bernstein D. Containers and cloud: From LXC to docker to kubernetes[J]. IEEE Cloud Computing, 2014, 1(3): 81-84.
- [6] Fenn M, Murphy M A, Martin J, et al. An evaluation of KVM for use in cloud computing[C/OL]//2nd International Conference on the Virtual Computing Initiative. [2021-01-10]. <https://people.cs.clemson.edu/~jmarty/papers/icvci08.pdf>.
- [7] Marathe N, Gandhi A, Shah J M. Docker swarm and kubernetes in cloud computing environment[C]//3rd International Conference on Trends in Electronics and Informatics. Piscataway, USA: IEEE, 2019. DOI: 10.1109/ICOEI.2019.8862654.
- [8] Georgakopoulos D, Jayaraman P P, Fazia M, et al. Internet of things and edge cloud computing roadmap for manufacturing[J]. IEEE Cloud

- Computing, 2016, 3(4): 66–73.
- [9] 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.
- [10] 东方财富网. 边缘计算(BK0860) [EB/OL]. [2021–01–16]. <http://quote.eastmoney.com/bk/90.BK0860.html>.
Easy money.com. Edge computing(BK0860) [EB/OL]. [2021–01–16]. <http://quote.eastmoney.com/bk/90.BK0860.html>.
- [11] Shaout A, Colella D, Awad S. Advanced driver assistance systems – past, present and future [C] //The Seventh International Computer Engineering Conference. Piscataway, USA: IEEE, 2011. DOI: 10.1109/ICENCO.2011.6153935.
- [12] Abbas N, Zhang Y, Taherkordi A, et al. Mobile edge computing: A survey [J]. IEEE Internet of Things Journal, 2018, 5(1): 450–465.
- [13] Mao Y Y, You C S, Zhang J, et al. A survey on mobile edge computing: The communication perspective [J]. IEEE Communications Surveys & Tutorials, 2017, 19(4): 2322–2358.
- [14] Satyanarayanan M, Bahl P, Caceres R, et al. The case for VM-based cloudlets in mobile computing [J]. IEEE Pervasive Computing, 2009, 8(4): 14–23.
- [15] Bonomi F, Milito R, Zhu J, et al. Fog computing and its role in the Internet of things [C] //The first edition of the MCC Workshop on Mobile Cloud Computing. New York, USA: ACM, 2012: 13–16.
- [16] Tran T X, Hajisami A, Pandey P, et al. Collaborative mobile edge computing in 5G networks: New paradigms, scenarios, and challenges [J]. IEEE Communications Magazine, 2017, 55(4): 54–61.
- [17] Sun Y H, Peng M G, Mao S W. Deep reinforcement learning-based mode selection and resource management for green fog radio access networks [J]. IEEE Internet of Things Journal, 2019, 6(2): 1960–1971.
- [18] 阿里云. 边缘节点服务 ENS [EB/OL]. [2020–12–22]. <https://help.aliyun.com/product/62684.html>.
Aliyun. Edge node service (ENS) [EB/OL]. [2020–12–22]. <https://help.aliyun.com/product/62684.html>.
- [19] Amazon Web Services. AWS for the Edge-Overview [EB/OL]. [2021–01–19]. <https://aws.amazon.com/cn/edge/>.
- [20] Ni L N, Zhang J Q, Jiang C J, et al. Resource allocation strategy in fog computing based on priced timed petri nets [J]. IEEE Internet of Things Journal, 2017, 4(5): 1216–1228.
- [21] Sato K, Azuma S I. Secure real-time control through fog computation [J]. IEEE Transactions on Industrial Informatics, 2019, 15(2): 1017–1026.
- [22] Li X C, Zhou Z B, Guo J Q, et al. Aggregated multi-attribute query processing in edge computing for industrial IoT applications [J]. Computer Networks, 2019, 151: 114–123.
- [23] Fraga-Lamas P, Lopez-Iturri P, Celaya-Echarri M, et al. Design and empirical validation of a Bluetooth 5 fog computing based industrial CPS architecture for intelligent Industry 4.0 shipyard workshops [J]. IEEE Access, 2020, 8: 45496–45511.
- [24] Fernández-Caramés T M, Fraga-Lamas P, Suárez-Albela M, et al. A fog computing based cyber-physical system for the automation of pipe-related tasks in the Industry 4.0 shipyard [J]. Sensors, 2018, 18(6). DOI: 10.3390/s18061961.
- [25] Blanco-Novoa O, Fernandez-Carames T M, Fraga-Lamas P, et al. A practical evaluation of commercial industrial augmented reality systems in an Industry 4.0 shipyard [J]. IEEE Access, 2018, 6: 8201–8218.
- [26] Rocha A D, Tripa J, Alemão D, et al. Agent-based plug and produce cyber-physical production system-test case [C] //IEEE 17th International Conference on Industrial Informatics. Piscataway, USA: IEEE, 2019. DOI: 10.1109/INDIN41052.2019.8972169.
- [27] Qiao H H, Wang T Y, Wang P. A tool wear monitoring and prediction system based on multiscale deep learning models and fog computing [J]. The International Journal of Advanced Manufacturing Technology, 2020, 108: 2367–2384.
- [28] 姚锡凡, 蓝宏宇, 陶韬, 等. 基于云雾结合的工件深度学习识别问题研究 [J]. 华南理工大学学报(自然科学版), 2019, 47(12): 1–8.
Yao X F, Lan H Y, Tao T, et al. Deep-learning recognition of workpieces based on cloud and fog computing [J]. Journal of South China University of Technology (Natural Science Edition), 2019, 47(12): 1–8.
- [29] Li L Z, Ota K, Dong M X. Deep learning for smart industry: Efficient manufacture inspection system with fog computing [J]. IEEE Transactions on Industrial Informatics, 2018, 14(10): 4665–4673.
- [30] Ghazal M, Basmaji T, Yaghi M, et al. Cloud-based monitoring of thermal anomalies in industrial environments using AI and the internet of robotic things [J]. Sensors, 2020, 20(21). DOI: 10.3390/s20216348.
- [31] Mocanu S, Geampalia G, Chenaru O, et al. Fog-based solution for real-time monitoring and data processing in manufacturing [C] //22nd International Conference on System Theory, Control and Computing. Piscataway, USA: IEEE, 2018. DOI: 10.1109/ICSTCC.2018.8540783.
- [32] Ashjaei M, Bengtsson M. Enhancing smart maintenance management using fog computing technology [C] //IEEE International Conference on Industrial Engineering and Engineering Management. Piscataway, USA: IEEE, 2017. DOI: 10.1109/IEEM.2017.8290155.
- [33] Tang H, Li D, Wan J F, et al. A reconfigurable method for intelligent manufacturing based on industrial cloud and edge intelligence [J]. IEEE Internet of Things Journal, 2020, 7(5): 4248–4259.
- [34] Liang F, Yu W, Liu X, et al. Toward edge-based deep learning in industrial Internet of Things [J]. IEEE Internet of Things Journal, 2020,

- 7(5): 4329–4341.
- [35] Sano T, Uruguchi T, Deguchi H, et al. Assignment of encryption data using fog computing technology in food manufacturing industry [C/OL] // 12th Asian Control Conference. Piscataway, USA: IEEE. (2019–07–18) [2021–04–08]. <https://ieeexplore.ieee.org/document/8765046>.
- [36] Zhao S, Li F H, Li H W, et al. Smart and practical privacy-preserving data aggregation for fog-based smart grids [J]. IEEE Transactions on Information Forensics and Security, 2020, 16: 521–536.
- [37] O'Donovan P, Gallagher C, Leahy K, et al. A comparison of fog and cloud computing cyber-physical interfaces for Industry 4.0 real-time embedded machine learning engineering applications [J]. Computers in Industry, 2019, 110: 12–35.
- [38] Zhao Z H, Lin P, Shen L D, et al. IoT edge computing-enabled collaborative tracking system for manufacturing resources in industrial park [J]. Advanced Engineering Informatics, 2020, 43. DOI: 10.1016/j.aei.2020.101044.
- [39] Lai C F, Chien W C, Yang L T, et al. LSTM and edge computing for big data feature recognition of industrial electrical equipment [J]. IEEE Transactions on Industrial Informatics, 2019, 15(4): 2469–2477.
- [40] Salhaoui M, Guerrero-González A, Arioua M, et al. Smart industrial IoT monitoring and control system based on UAV and cloud computing applied to a concrete plant [J]. Sensors, 2019, 19(15). DOI: 10.3390/s19153316.
- [41] Peng M G, Quek T Q S, Mao G Q, et al. Artificial-intelligence-driven fog radio access networks: Recent advances and future trends [J]. IEEE Wireless Communications, 2020, 27(2): 12–13.
- [42] Kaneko Y, Yokoyama Y, Monma N, et al. A microservice-based industrial control system architecture using cloud and MEC [M] // Lecture Notes in Computer Science, vol. 12407. Berlin, Germany: Springer, 2020: 18–32.
- [43] Wang T, Ke H X, Zheng X, et al. Big data cleaning based on mobile edge computing in industrial sensor-cloud [J]. IEEE Transactions on Industrial Informatics, 2020, 16(2): 1321–1329.
- [44] Zhu M L, Liu C. A correlation driven approach with edge services for predictive industrial maintenance [J]. Sensors, 2018, 18(6). DOI: 10.3390/s18061844.
- [45] Zhang J, Deng C Y, Zheng P, et al. Development of an edge computing-based cyber-physical machine tool [J]. Robotics and Computer-Integrated Manufacturing, 2021, 67. DOI: 10.1016/j.rcim.2020.102042.
- [46] 董锴文, 孙彦景, 陈岩, 等. 矿山生产作业场景视频结构化分析关键技术 [J/OL]. 煤炭学报. (2021–03–17) [2021–04–18]. <https://doi.org/10.13225/j.cnki.jccs.2020.2019>.
Dong K W, Sun Y J, Chen Y, et al. Key technology of video structured analysis in mine production operation scenario [J/OL]. Journal of China Coal Society. (2021–03–17) [2021–04–18]. <https://doi.org/10.13225/j.cnki.jccs.2020.2019>.
- [47] 马靖, 王译晨, 赵明, 等. 基于数字孪生的生产单元可视化管控方法 [J/OL]. 计算机集成制造系统. (2021–03–24) [2021–04–08]. <http://kns.cnki.net/kcms/detail/11.5946.TP.20210323.1028.006.html>.
Ma J, Wang Z C, Zhao M, et al. Visualization management and control methods of production cell based on digital twin [J/OL]. Computer Integrated Manufacturing Systems. (2021–03–24) [2021–04–08]. <http://kns.cnki.net/kcms/detail/11.5946.TP.20210323.1028.006.html>.
- [48] 于天琪, 胡剑凌, 金炯, 等. 基于移动边缘计算的车载 CAN 网络入侵检测方法 [J]. 计算机科学, 2021, 48(1): 34–39.
Yu T Q, Hu J L, Jin J, et al. Mobile edge computing based in-vehicle CAN network intrusion detection method [J]. Computer Science, 2021, 48(1): 34–39.
- [49] 彭超, 唐向红, 陆见光. 基于边缘计算的轴承故障诊断 [J]. 组合机床与自动化加工技术, 2020, (12): 52–55.
Peng C, Tang X H, Lu J G. Bearing fault diagnosis based on edge computing [J]. Modular Machine Tool & Automatic Manufacturing Technique, 2020, (12): 52–55.
- [50] 网宿科技股份有限公司. 基于边缘计算的 AI 表面无人质检系统解决方案 [J]. 自动化博览, 2020(10): 58–61.
Wangsu Science and Technology Co., Ltd. AI surface unmanned quality inspection system solution based on edge computing [J]. Automation Panorama, 2020(10): 58–61.
- [51] 李兴, 侯伟, 杨宇, 等. 论电力基建业务中边缘计算的应用与实践 [J]. 电气工程学报, 2020, 15(4): 153–160.
Li X, Hou W, Yang Y, et al. On the application and practice of edge computing in infrastructure [J]. Journal of Electrical Engineering, 2020, 15(4): 153–160.
- [52] 胡金磊, 朱泽锋, 林孝斌, 等. 变电站无人机巡检边缘计算框架设计及资源调度方法 [J]. 高电压技术, 2021, 47(2): 425–433.
Hu J L, Zhu Z F, Lin X B, et al. Framework design and resource scheduling method for edge computing in substation UAV inspection [J]. High Voltage Engineering, 2021, 47(2): 425–433.
- [53] 刘曼伊, 臧永立, 朱超平. 基于边缘计算的铁路无人值守站监测应用研究 [J]. 计算机仿真, 2021, 38(3): 51–55.
Liu Y Y, Zang Y L, Zhu C P. Research on monitoring application of railway unattended station based on edge computing [J]. Computer Simulation, 2021, 38(3): 51–55.
- [54] 陈和恩, 何汉武, 吴悦明. VR/AR 与智能制造: 应用领域、核心技术 [J]. 机电工程技术, 2021, 50(2): 1–4, 18.

- Chen H E, He H W, Wu Y M. VR/AR and intelligent manufacturing system: Application areas, core technologies [J]. *Mechanical & Electrical Engineering Technology*, 2021, 50(2): 1–4, 18.
- [55] Fernández-Caramés T M, Fraga-Lamas P, Suárez-Albela M, et al. A fog computing and cloudlet based augmented reality system for the Industry 4.0 shipyard [J]. *Sensors*, 2018, 18(6). DOI: 10.3390/s18061798.
- [56] 陶飞, 刘蔚然, 刘检华, 等. 数字孪生及其应用探索 [J]. *计算机集成制造系统*, 2018, 24(1): 1–18.
- Tao F, Liu W R, Liu J H, et al. Digital twin and its potential application exploration [J]. *Computer Integrated Manufacturing Systems*, 2018, 24(1): 1–18.
- [57] 王大伟, 王卓, 王鹏, 等. 基于边缘计算的云原生机器人系统 [J]. *智能科学与技术学报*, 2020, 2(3): 275–283.
- Wang D W, Wang Z, Wang P, et al. Cloud native robot system based on edge computing [J]. *Chinese Journal of Intelligent Science and Technology*, 2020, 2(3): 275–283.
- [58] 丁鹏, 薛裕颖, 熊小敏, 等. 5G + 边缘计算在智能汽车柔性制造中的应用 [J]. *电子技术应用*, 2020, 46(12): 26–31.
- Ding P, Xue Y Y, Xiong X M, et al. Application of 5G + edge computing in flexible manufacturing of intelligent vehicles [J]. *Application of Electronic Technique*, 2020, 46(12): 26–31.
- [59] 郑松, 郑蓉. 工业互联网连接与安全模型研究 [J]. *自动化博览*, 2021, 38(1): 72–77.
- Zheng S, Zheng R. Research on connection and security model of industrial internet [J]. *Automation Panorama*, 2021, 38(1): 72–77.
- [60] Giehl A, Schneider P, Busch M, et al. Edge-computing enhanced privacy protection for industrial ecosystems in the context of SMEs [C] // 12th CMI Conference on Cybersecurity and Privacy. Piscataway, USA: IEEE, 2019. DOI: 10.1109/CMI48017.2019.8962138.
- [61] Zhang Y S, Wang P, Huang H, et al. Privacy-assured FogCS: Chaotic compressive sensing for secure industrial big image data processing in fog computing [J]. *IEEE Transactions on Industrial Informatics*, 2021, 17(5): 3401–3411.
- [62] Usman M, Jolfaei A, Jan M A. RaSEC: An intelligent framework for reliable and secure multilevel edge computing in industrial environments [J]. *IEEE Transactions on Industry Applications*, 2020, 56(4): 4543–4551.
- [63] McMahan B, Moore E, Ramage D, et al. Communication-efficient learning of deep networks from decentralized data [C] // 20th International Conference on Artificial Intelligence and Statistics. Cambridge, USA: MIT Press, 2017, 54: 273–1282.
- [64] Lu Y L, Huang X H, Dai Y Y, et al. Differentially private asynchronous federated learning for mobile edge computing in urban informatics [J]. *IEEE Transactions on Industrial Informatics*, 2020, 16(3): 2134–2143.
- [65] Usman M, Jan M A, Jolfaei A, et al. A distributed and anonymous data collection framework based on multilevel edge computing architecture [J]. *IEEE Transactions on Industrial Informatics*, 2020, 16(9): 6114–6123.
- [66] Zhang Y S, Huang H, Yang L X, et al. Serious challenges and potential solutions for the industrial Internet of Things with edge intelligence [J]. *IEEE Network*, 2019, 33(5): 41–45.
- [67] Li C X, Liao X F, Jin H. Enhancing application performance via DAG-driven scheduling in task parallelism for cloud center [J]. *Peer-to-Peer Networking and Applications*, 2019, 12: 381–391.
- [68] Li S H, Chen W H, Chen Y F, et al. Makespan-minimized computation offloading for smart toys in edge-cloud computing [J]. *Electronic Commerce Research and Applications*, 2019, 37. DOI: 10.1016/j.elerap.2019.100884.
- [69] Mazouzi H, Achir N, Boussetta K. Elastic offloading of multitasking applications to mobile edge computing [C] // 22nd International ACM Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems. New York, USA: ACM, 2019: 307–314.
- [70] Liu H L, Cao L, Pei T R, et al. A fast algorithm for energy-saving offloading with reliability and latency requirements in multi-access edge computing [J]. *IEEE Access*, 2019, 8: 151–161.
- [71] Ren J K, Yu G D, Cai Y L, et al. Latency optimization for resource allocation in mobile-edge computation offloading [J]. *IEEE Transactions on Wireless Communications*, 2018, 17(8): 5506–5519.
- [72] Yousefpour A, Ishigaki G, Gour R, et al. On reducing IoT service delay via fog offloading [J]. *IEEE Internet of Things Journal*, 2018, 5(2): 998–1010.
- [73] Mukherjee M, Kumar S, Mavromoustakis C X, et al. Latency-driven parallel task data offloading in fog computing networks for industrial applications [J]. *IEEE Transactions on Industrial Informatics*, 2020, 16(9): 6050–6058.
- [74] Lin R P, Zhou Z J, Luo S, et al. Distributed optimization for computation offloading in edge computing [J]. *IEEE Transactions on Wireless Communications*, 2020, 19(12): 8179–8194.
- [75] Chiti F, Fantacci R, Picano B. A matching theory framework for tasks offloading in fog computing for IoT systems [J]. *IEEE Internet of Things Journal*, 2018, 5(6): 5089–5096.
- [76] Jiang Y L, Chen Y S, Yang S W, et al. Energy-efficient task offloading for time-sensitive applications in fog computing [J]. *IEEE Systems Journal*, 2019, 13(3): 2930–2941.
- [77] Liu C F, Bennis M, Debbah M, et al. Dynamic task offloading and resource allocation for ultra-reliable low-latency edge computing [J]. *IEEE Transactions on Communications*, 2019, 67(6): 4132–4150.

- [78] Wu Z K, Li B, Fei Z S, et al. Energy-efficient robust computation offloading for fog-IoT systems [J]. IEEE Transactions on Vehicular Technology, 2020, 69(4): 4417–4425.
- [79] Baranwal G, Vidyarthi D P. Computation offloading model for smart factory [J]. Journal of Ambient Intelligence and Humanized Computing, 2020. DOI: 10.1007/s12652-020-02564-0.
- [80] Zhu Q L, Si B J, Yang F F, et al. Task offloading decision in fog computing system [J]. China Communications, 2017, 14(11): 59–68.
- [81] Huang L, Feng X, Zhang C, et al. Deep reinforcement learning-based joint task offloading and bandwidth allocation for multi-user mobile edge computing [J]. Digital Communications and Networks, 2019, 5(1): 10–17.
- [82] Liu L Q, Chang Z, Guo X J, et al. Multiobjective optimization for computation offloading in fog computing [J]. IEEE Internet of Things Journal, 2018, 5(1): 283–294.
- [83] 姚维, 黄俊, 移动边缘计算中的无人机群协同任务卸载策略研究 [J/OL]. 重庆邮电大学学报(自然科学版). (2021-03-17) [2021-04-17]. <http://kns.cnki.net/kcms/detail/50.1181.N.20210317.0908.002.html>.
Yao W, Huang J. Research on UAVs cooperative task offloading strategy in mobile edge computing [J/OL]. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition). (2021-03-17) [2021-04-17]. <http://kns.cnki.net/kcms/detail/50.1181.N.20210317.0908.002.html>.
- [84] Teerapittayanon S, McDanel B, Kung H T. BranchyNet: Fast inference via early exiting from deep neural networks [C] // 23rd International Conference on Pattern Recognition. Piscataway, USA: IEEE, 2016. DOI: 10.1109/ICPR.2016.7900006.
- [85] Wang Q Z, Wang K, Li Q, et al. MBNN: A multi-branch neural network capable of utilizing industrial sample unbalance for fast inference [J]. IEEE Sensors Journal, 2021, 21(2): 1809–1819.
- [86] Li L Z, Ota K, Dong M X. Deep learning for smart industry: Efficient manufacture inspection system with fog computing [J]. IEEE Transactions on Industrial Informatics, 2018, 14(10): 4665–4673.
- [87] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [C/OL] // 3rd International Conference on Learning Representations. San Diego, USA: ICLR, 2015. [2021-04-06]. <https://arxiv.org/pdf/1409.1556v6.pdf>.
- [88] Teerapittayanon S, McDanel B, Kung H T. Distributed deep neural networks over the cloud, the edge and end devices [C] // 37th International Conference on Distributed Computing Systems. Piscataway, USA: IEEE, 2017. DOI: 10.1109/ICDCS.2017.226.
- [89] Li H S, Hu C H, Jiang J Y, et al. JALAD: Joint accuracy-and latency-aware deep structure decoupling for edge-cloud execution [C] // 24th International Conference on Parallel and Distributed Systems. Piscataway, USA: IEEE, 2018. DOI: 10.1109/PADSW.2018.8645013.
- [90] Velasco-Montero D, Fernandez-Berni J, Carmona-Galan R, et al. Optimum selection of DNN model and framework for edge inference [J]. IEEE Access, 2018, 6: 51680–51692.
- [91] 朱禹涛, 肖霖, 陈泽仁, 等. 面向智慧工厂的工业互联网边缘智能协同计算技术研究 [J]. 信息通信技术与政策, 2021, 47(3): 1–5.
Zhu Y T, Xiao L, Chen Z R, et al. Research on edge intelligent cooperative computing technology in industrial Internet for intelligent factory [J]. Information and Communications Technology and Policy, 2021, 47(3): 1–5.
- [92] Xu Z C, Liang W F, Xu W Z, et al. Efficient algorithms for capacitated cloudlet placements [J]. IEEE Transactions on Parallel and Distributed Systems, 2016, 27(10): 2866–2880.
- [93] Ren Y Z, Zeng F, Li W J, et al. A low-cost edge server placement strategy in wireless metropolitan area networks [C] // 27th International Conference on Computer Communication and Networks. Piscataway, USA: IEEE, 2018. DOI: 10.1109/ICCCN.2018.8487438.
- [94] Lin C C, Yang J W. Cost-efficient deployment of fog computing systems at logistics centers in Industry 4.0 [J]. IEEE Transactions on Industrial Informatics, 2018, 14(10): 4603–4611.
- [95] 刘伯阳, 杨宁乐, 马杰, 等. 无人机移动边缘计算位置部署与资源分配方案 [J]. 西安邮电大学学报, 2021, 26(1): 33–38, 59.
Liu B Y, Yang N L, Ma J, et al. Optimal deployment and resource allocation for UAV-enabled mobile edge computing networks [J]. Journal of Xi'an University of Posts and Telecommunications, 2021, 26(1): 33–38, 59.
- [96] Sun W, Liu J J, Yue Y L, et al. Double auction-based resource allocation for mobile edge computing in industrial Internet of Things [J]. IEEE Transactions on Industrial Informatics, 2018, 14(10): 4692–4701.
- [97] Jin A L, Song W, Zhuang W H. Auction-based resource allocation for sharing cloudlets in mobile cloud computing [J]. IEEE Transactions on Emerging Topics in Computing, 2018, 6(1): 45–57.
- [98] Sun Y, Zhang N. A resource-sharing model based on a repeated game in fog computing [J]. Saudi Journal of Biological Sciences, 2017, 24(3): 687–694.
- [99] Taneja M, Davy A. Resource aware placement of IoT application modules in fog-cloud computing paradigm [C] // IFIP/IEEE Symposium on Integrated Network and Service Management. Piscataway, USA: IEEE, 2017. DOI: 10.23919/INM.2017.7987464.
- [100] Tang H L, Li C L, Bai J P, et al. Dynamic resource allocation strategy for latency-critical and computation-intensive applications in cloud-edge environment [J]. Computer Communications, 2019, 134: 70–82.
- [101] Zhao L, Wang J D, Liu J J, et al. Optimal edge resource allocation in IoT-based smart cities [J]. IEEE Network, 2019, 33(2): 30–35.

- [102] Hong C H, Lee K, Kang M, et al. qCon: QoS-aware network resource management for fog computing [J]. *Sensors*, 2018, 18(10). DOI: 10.3390/s18103444.
- [103] Afrin M, Jin J, Rahman A, et al. Multi-objective resource allocation for edge cloud based robotic workflow in smart factory [J]. *Future Generation Computer Systems*, 2019, 97: 119–130.
- [104] 张雷, 李琳, 陈鸿龙, 等. 一种面向工业边缘计算应用的缓存替换算法 [J/OL]. *计算机研究与发展*. (2021-02-25) [2021-03-29]. <http://kns.cnki.net/kcms/detail/11.1777.TP.20210225.1045.002.html>.
Zhang L, Li L, Chen H L, et al. A cache replacement algorithm for industrial edge computing application [J/OL]. *Journal of Computer Research and Development*. (2021-02-25) [2021-03-29]. <http://kns.cnki.net/kcms/detail/11.1777.TP.20210225.1045.002.html>.
- [105] Melcher J. Design and implementation of a container-based architecture for real-time control applications [D]. Stuttgart, Germany: University of Stuttgart, 2018.
- [106] 王硕, 郑爱云, 刘怀煜. 应用于智能制造的边缘计算任务调度算法研究 [J]. *制造业自动化*, 2020, 42(12): 98–105.
Wang S, Zheng A Y, Liu H Y. Task scheduling strategy of edge computing system in intelligent manufacturing [J]. *Manufacturing Automation*, 2020, 42(12): 98–105.
- [107] Li X, Liu Y M, Ji H, et al. Optimizing resources allocation for fog computing-based internet of things networks [J]. *IEEE Access*, 2019, 7: 64907–64922.
- [108] 杨乐, 李萌, 叶欣宇, 等. 融合边缘计算与区块链的工业互联网资源优化配置研究 [J]. *高技术通讯*, 2020, 30(12): 1253–1263.
Yang L, Li M, Ye X Y, et al. Research on resource optimization and allocation for industrial internet based on edge computing and blockchain [J]. *Chinese High Technology Letters*, 2020, 30(12): 1253–1263.
- [109] Xu F M, Yang F, Bao S J, et al. DQN inspired joint computing and caching resource allocation approach for software defined information-centric Internet of Things network [J]. *IEEE Access*, 2019, 7: 61987–61996.
- [110] Pham Q V, Leanh T, Tran N H, et al. Decentralized computation offloading and resource allocation for mobile-edge computing: A matching game approach [J]. *IEEE Access*, 2018, 6: 75868–75885.
- [111] Zeng D Z, Gu L, Pan S L, et al. Resource management at the network edge: A deep reinforcement learning approach [J]. *IEEE Network*, 2019, 33(3): 26–33.
- [112] Avgeris M, Dechouniotis D, Athanasopoulos N, et al. Adaptive resource allocation for computation offloading: A control-theoretic approach [J]. *ACM Transactions on Internet Technology*, 2019, 19(2). DOI: 10.1145/3284553.
- [113] Tseng F H, Tsai M S, Tseng C W, et al. A lightweight autoscaling mechanism for fog computing in industrial applications [J]. *IEEE Transactions on Industrial Informatics*, 2018, 14(10): 4529–4537.
- [114] Wang W, Zhao Y L, Tornatore M, et al. Virtual machine placement and workload assignment for mobile edge computing [C] // 6th International Conference on Cloud Networking. Piscataway, USA: IEEE, 2017. DOI: 10.1109/CloudNet.2017.8071527.
- [115] Ruan L, Liu Z B, Qiu X S, et al. Resource allocation and distributed uplink offloading mechanism in fog environment [J]. *Journal of Communications and Networks*, 2018, 20(3): 247–256.
- [116] Zhang J, Xia W W, Yan F, et al. Joint computation offloading and resource allocation optimization in heterogeneous networks with mobile edge computing [J]. *IEEE Access*, 2018, 6: 19324–19337.
- [117] Li S C, Zhang N, Lin S Y, et al. Joint admission control and resource allocation in edge computing for Internet of Things [J]. *IEEE Network*, 2018, 32(1): 72–79.
- [118] 苏命峰, 王国军, 李仁发. 边云协同计算中基于预测的资源部署与任务调度优化 [J/OL]. *计算机研究与发展*. (2021-03-16) [2021-04-20]. <http://kns.cnki.net/kcms/detail/11.1777.tp.20210316.1150.004.html>.
Su M F, Wang G J, Li R F. Resource deployment with prediction and task scheduling optimization in edge cloud collaborative computing [J/OL]. *Journal of Computer Research and Development*. (2021-03-16) [2021-04-20]. <http://kns.cnki.net/kcms/detail/11.1777.tp.20210316.1150.004.html>.
- [119] Aazam M, Harras K A, Zeadally S. Fog computing for 5G tactile industrial Internet of Things: QoE-aware resource allocation model [J]. *IEEE Transactions on Industrial Informatics*, 2019, 15(5): 3085–3092.
- [120] Zhu Z W, Wu F, Cao J, et al. A thread-oriented memory resource management framework for mobile edge computing [J]. *IEEE Access*, 2019, 7: 45881–45890.
- [121] Ha K, Pillai P, Richter W, et al. Just-in-time provisioning for cyber foraging [C] // 11th Annual International Conference on Mobile Systems, Applications, and Services. New York, USA: ACM, 2013: 153–166.
- [122] Lu W, Meng X Y, Guo G F. Fast service migration method based on virtual machine technology for MEC [J]. *IEEE Internet of Things Journal*, 2019, 6(3): 4344–4354.
- [123] Zhou C J, Chen H. Development of a PLC virtual machine orienting IEC 61131-3 standard [C] // International Conference on Measuring Technology and Mechatronics Automation. Piscataway, USA: IEEE, 2009. DOI: 10.1109/ICMTMA.2009.422.

- [124] Breivold H P, Sandström K. Virtualize for test environment in industrial automation [C] //2014 IEEE Emerging Technology and Factory Automation. Piscataway, USA: IEEE, 2014. DOI: 10.1109/ETFA.2014.7005089.
- [125] Elgazar A, Harras K. Teddybear: Enabling efficient seamless container migration in user-owned edge platforms [C] //IEEE International Conference on Cloud Computing Technology and Science. Piscataway, USA: IEEE, 2019. DOI: 10.1109/CloudCom.2019.00022.
- [126] Avino G, Malinverno M, Malandrino F, et al. Characterizing docker overhead in mobile edge computing scenarios [C] //the Workshop on Hot Topics in Container Networking and Networked Systems. New York, USA: ACM, 2017: 30–35.
- [127] Ismail B I, Goortani E M, Ab Karim M B, et al. Evaluation of docker as edge computing platform [C] //IEEE Conference on Open Systems. Piscataway, USA: IEEE, 2015. DOI: 10.1109/ICOS.2015.7377291.
- [128] Bellavista P, Zanni A. Feasibility of fog computing deployment based on docker containerization over raspberrypi [C] //Proceedings of the 18th International Conference on Distributed Computing and Networking. New York, USA: ACM, 2017, 16: 1–10.
- [129] Yokogawa selected by ExxonMobil as the open process automation (OPA) system integrator [EB/OL]. (2019–07–10) [2019–08–28]. <https://www.controlglobal.com/industrynews/2019/yokogawa-selected-by-exxonmobil-as-the-open-process-automation-opa-system-integrator/>.
- [130] Höhr S, Tasci T, Verl A. Realization of data analytics projects in manufacturing using a microservice-based approach [C] //IEEE International Conference on Mechatronics. Piscataway, USA: IEEE, 2019. DOI: 10.1109/ICMECH.2019.8722926.
- [131] Goldschmidt T, Hauck-Stattelmann S, Malakuti S, et al. Container-based architecture for flexible industrial control applications [J]. Journal of Systems Architecture, 2018, 84: 28–36.
- [132] Google. Kubernetes [EB/OL]. [2021–01–14]. <https://kubernetes.io/>.
- [133] Docker Documentation. Docker swarm [EB/OL]. [2021–01–14]. <https://docs.docker.com/engine/reference/commandline/swarm/>.
- [134] Rancher. K3s: Lightweight Kubernetes [EB/OL]. [2019–10–22]. <https://k3s.io/>.
- [135] HUAWEI. KubeEdge [CP/OL]. [2019–10–22]. <https://kubedge.io/en/>.
- [136] Messinger D, Lewis G. Application virtualization as a strategy for cyber foraging in resource-constrained environments [Z/OL]. [2020–01–02]. <https://semanticscholar.org/>.
- [137] Wu S, Mei C, Jin H, et al. Android unikernel: Gearing mobile code offloading towards edge computing [J]. Future Generation Computer Systems, 2018, 86: 694–703.
- [138] Ren L, Sun Y Q, Cui J, et al. Bearing remaining useful life prediction based on deep autoencoder and deep neural networks [J]. Journal of Manufacturing Systems, 2018, 48(Part C): 71–77.
- [139] Li X, Zhang W, Ding Q. Deep learning-based remaining useful life estimation of bearings using multi-scale feature extraction [J]. Reliability Engineering & System Safety, 2019, 182: 208–218.
- [140] Hong C H, Spence I, Nikolopoulos D S. GPU virtualization and scheduling methods: A comprehensive survey [J]. ACM Computing Surveys, 2017, 50(3). DOI: 10.1145/3068281.
- [141] Dowty M, Sugerman J. GPU virtualization on VMware's hosted I/O architecture [J]. ACM SIGOPS Operating Systems Review, 2009, 43(3). DOI: 10.1145/1618525.1618534.
- [142] NVIDIA. Nvidia-docker: Build and run Docker containers leveraging NVIDIA GPUs [EB/OL]. [2019–11–27]. <https://github.com/NVIDIA/nvidia-docker>.
- [143] Bag G, Lednicki L, Landernäs K, et al. Experiments on approaches of virtualization for industrial Internet of Things applications [C] //24th IEEE International Conference on Emerging Technologies and Factory Automation. Piscataway, USA: IEEE, 2019. DOI: 10.1109/ETFA.2019.8868971.
- [144] Abeni L, Balsini A, Cucinotta T. Container-based real-time scheduling in the Linux kernel [J]. ACM SIGBED Review, 2019, 16(3). DOI: 10.1145/3373400.3373405.

作者简介

王其朝(1994–), 男, 博士生. 研究领域为边缘计算, 工业互联网, 边缘智能.

金光淑(1979–), 女, 硕士. 研究领域为工业云平台, 工业通信.

李 庆(1992–), 女, 硕士生. 研究领域为工业控制网络.