

工业算力网络的挑战 and 关键技术

Challenges and Crucial Technologies of Industrial Computing-First Network

北京邮电大学 许士勋, 张奥, 宋闻萱, 黄潇洁, 许方敏, 赵成林

摘要: 工业算力网络将算力网络引入到工业互联网中, 实现对工业场景下算力和网络资源的统一编排调度, 提高了资源的利用率, 满足了工业互联网智能化发展趋势下低时延、高可靠的业务需求。本文在介绍工业算力网络架构的基础上, 结合工业任务及工业互联网特性, 对算力网络赋能工业互联网面临的挑战进行了分析, 并对实现工业算力网络的关键技术进行了总结和构想。

关键词: 工业互联网; 算力网络

Abstract: The Industrial Computing First Network integrates Computing First Network into the Industrial Internet to realize the unified arrangement and scheduling of computing power and network resources in industrial scenarios, thereby improving the utilization rate of resources. This facilitates meeting the service requirements of low delay and high reliability in line with the intelligent development trend of the Industrial Internet. Based on the introduction of the architecture of Industrial Computing First Network, this paper analyzes the challenges faced by Computing First Network to enable Industrial Internet, and summarizes and proposes the key technologies for realizing Industrial Computing First Network, taking the characteristics of industrial tasks and Industrial Internet into consideration.

Key words: Industrial internet; Computing-first network

1 引言

工业互联网作为新一代网络信息技术和制造业的融合产物, 是实现产业数字化、网络化和智能化发展的重要途径, 也是实现工业环境下人、机、物互联的重要基础设施。随着工业互联网中智能应用的部署, 其对网络时延、可靠性及计算能力的要求不断提高。目前, 工业互联网中泛在的算力及网络资源利用率低, 难以满足工业应用的业务需求。

算力网络是算网融合演进的新型网络架构, 通过对计算和网络资源统一编排调度, 可以实现两者利用率最大化。通过算力网络对工业互联网赋能, 构建工业算力网络, 并进行云网边端等计算、存储、网络资源的协同调度, 成为满足工业互联网智能业务需求的重要方式。

目前, 工业算力网络的发展尚处于起步阶段, 演进路线和技术体系初具雏形, 亟需产业界上下游联合推进, 以在算力度量、算力标识、算力感知、算力路由、算力调度、算力交易等技术方向实现突破。本文将从工业算力网络架构面临的挑战和其中的关键技术等方面进行逐一介绍。

2 工业算力网络架构

工业算力网络以工业互联网为基础, 通过算力网络进行资源配给实现工业互联网的升级。从逻辑功能上工业算力网络可以划分为工业设备接入层、算力网络资源层、算力网络控制层和算力网络应用服务层, 如图1所示。

其中工业设备接入层包括各种工业现场设备, 对算力和网络需求较高, 同时也进行算力执行。工业算力网络资源层包括由算力和网络基础设施构成的算力及网络资源。算力网络控制层对资源层中的异构算力资源进行统一度量建模, 并汇聚算力网络资源信息, 同时对用户业务需求进行分析解构和标识, 随后根据业务需求对全网资源进行协同调度, 实现服务灵活动态部署, 为用户业务选择合适的转发路径及算力节点。工业算力网络应用服务层包含算力交易平台及为用户提供各类应用服务的可编程接口, 用户经由工业算力服务层向算力交易平

台发起算力请求达成算力交易，调用可编程接口进行资源调用完成应用服务。

如利用卷积神经网络等方式对产品图像进行处理来实现划痕识别等质量检测，当设备产生算力需求时，向算力交易平台发起算力请求达成算力交易，并调用应用服务接口将业务信息传递到控制层完成对业务的度量、解构和算力网络资源的调度，将业务数据经合适路径路由至算力执行层中的合适节点进行处理，最终返回处理结果完成算力请求。

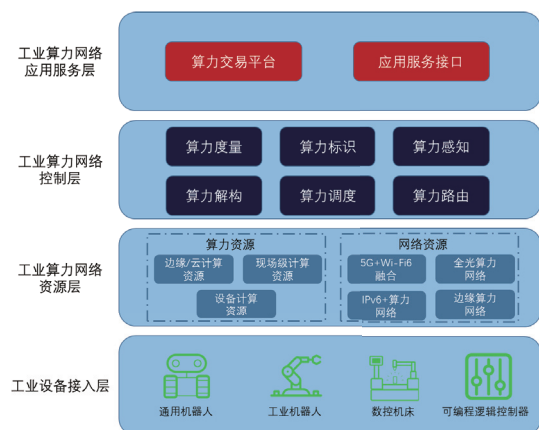


图1 工业算力网络架构

3 工业算力网络面临的挑战

由于工业任务需要满足大带宽、低时延、低抖动、高可靠性、高安全性的特殊需求^[1]，以及工业网络带来的内外网割裂、生产设备类型及数量繁多、工业链条长、数据量大的特性^[2]，使得使用算力网络对工业互联网赋能存在诸多限制。

3.1 算力度量与标识挑战

实现算力与网络的深度融合的前提是对算力网络中的算力资源与网络资源进行统一的度量和标识，以便进行后续资源的关联与分配。

在工业算力网络中，不同设备搭载的芯片以及工厂内外网所使用的异构算力资源如CPU、GPU、FPGA、TPU和ASIC等不同类型的硬件差异性大，需要进行性能上的统一度量。对于度量不同的运算类型，如逻辑运算、并行计算、神经网络计算，同一硬件的运算能力也难以统一。而且在算网融合场景中，算力的度量还与服务节点的通信能力与服务能力息息相关。因此，如何将

网络、计算、存储资源与服务共同联合建模，是工业算网度量体系面临的一大挑战。

针对专业性高、差异性大的工厂内智能任务，对算力资源与网络资源进行统一且可验证的标识可为算力调度提供基础。工业外网中算力的流通依赖于对不同数据中心算力标识的统一，且工业计算任务的保密性高，对于算网中存在的大量个人算力资源，亟需统一的标识实现算力的可信鉴权与溯源。

3.2 算力感知挑战

算力感知是对算力网络中计算资源的位置、总量、使用量等状态信息和算力业务需求的实时感知，是解决多算力节点协同问题、计算任务与节点匹配问题的基础。

在基础网络中，感知主要集中在通信网络的环境变化，粒度差异小。而在工业算力网络中，不同类型算力，如硬件资源与服务资源，二者的状态变更频率差异巨大，缺少具有针对性的算力感知机制。现有的算力感知技术缺少预测机制，面对多变、突发的海量工业任务需求，难以为实时感知算网中计算资源与网络资源的状态提供技术支撑。而且目前算网中仍存在大量尚待发掘的算力资源，需设计对应的算力发现技术与算力全网通告技术为算力感知提供信息基础。

3.3 算力路由挑战

算力路由是算力网络的核心，它负责根据业务的需求，将其调度至最佳的节点进行处理，为用户提供高效的服务。传统网络路由仅根据网络状态信息进行调度，而算力路由实现的是网络、存储、计算的共同调度。

工业算网中算力资源分布于工厂内网和外网，目前尚缺少工业内外网多层次算网节点的路由方案。SRv6源路由技术依靠可编程性与兼容性，已在IPv6网络中广泛使用，如何使其满足工业算力网络的需求，是对工业算力路由提出的新挑战。

部分要求实时处理工业任务对算网路由的实时性提出了挑战，有些短时工业任务导致算力网络中的状态信息变化很快，如何平衡路由的准确性与其对状态信息的变化敏感度也是一大难题。随着算力网络规模增大，各部分网络的环境、状态不同，网络波动、堵塞等问题难以避免，这对工业算网中算力路由的稳定性无疑也是一个挑战。

3.4 算力调度与资源分配挑战

当使用算力网络对工业互联网进行赋能时，算力

资源对用户侧的呈现形式由端到端变成了泛在部署^[2],这使得在为业务进行算力调度和资源分配时,会面临更多的可变性和不确定性,即不同位置的资源由于到用户的距离不同、网络连接好坏不同、各节点的报价不同、计算、存储等多维资源的配比不同等而变得不再平等。

此外,由于工业互联网场景下的业务需求弹性较大,当业务需求激增,泛在部署的节点算力资源不能满足激增的业务需求,这意味着需要将用户业务需求进行划分,调度至多个算力节点进行处理。

如何根据不同用户业务的需求,选择合适的算力节点和调度路径,实现用户体验的一致性和算力资源的统一编排及高效利用变得更加复杂。

3.5 算力交易与安全问题挑战

当用户与算力节点进行算力交易时,最为关注的则是定价的合理性和隐私数据的安全性。算力交易双方往往由于所处算力级别或所属算力网络不同,而具有不同的定价规则,且相互间业务的适配性出现差异^[3],这使得定价的合理性很难得到双方的一致肯定,交易就很难达成。

另一方面的安全性,则是由于业务数据中包含的大量敏感的隐私数据,或算力请求者和提供服务的算力节点归属不同,使得算力请求者对业务数据有保密需求,而数据在进行传输过程中存在被攻击的风险,如遭受窃听,因此算力网络能否解决数据的安全问题是算力交易能否顺利达成的另一重要因素。

4 工业场景中算力网络的关键技术

4.1 算力度量、算力标识与算力感知技术

对算力规模不同及类型异构的算力节点,可以采用度量函数将异构算力归一化映射为浮点运算能力,并选取合理阈值将算力划分为不同等级^[4]。为了方便与用户需求进行匹配,可以同样采用浮点运算次数对用户业务算力使用量进行度量^[5]。此外,工业应用需求一种或多种特定运算能力,这意味着不同异构算力节点擅长不同业务,可以采用人工智能算法根据算力节点的异构算力组成,对算力类型进行隶属度估计^[5],以将算力节点与其擅长的用户业务进行匹配。

利用以上算力度量结果,结合算力节点的计算、通信属性,以及算力服务多样性,可以构成对算力资源的唯一可扩展标识^[6],用于对全网算力节点的管理和对用户需求的初步匹配。

完成了对算力资源的抽象表示,还需要构建全网算力资源的感知图模型,以进行资源的统一编排管理。根据信息收集的方式可以将感知分为:分布式层次化感知^[3]和结合SDN技术的集中式感知^[5,7]。由于集中式方案的安全性及可扩展性较差,可以采用分布式方案。在分布式层次化感知中,如图2所示,算力节点根据算力度量值划分的算力等级进行多层次自治组网,并将算力信息在各级自治组织网内进行全网通告,确保算力信息的实时更新;各级算力关键路由将算力信息进行汇聚,并共同维护一张算力信息更新度最大化的层次化算力网络感知图。

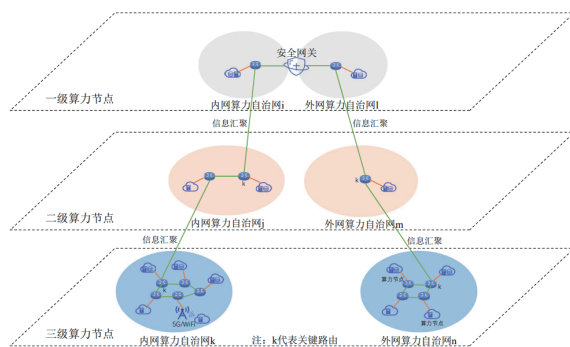


图2 分布式层次化算力感知

4.2 算力网络确定性传输技术

为了满足工业智能应用或控制类应用的确定性、可靠性需求,需要在工业算力网络中使用确定性传输技术。确定性传输技术可以实现一定界限内的时延、带宽、抖动、丢失率、可靠性要求下的传输,是一种保障业务流端到端传输时延相关服务质量的技术。

在算力资源及负载信息实时感知的基础上,将时间敏感网络(Time Sensitive Network, TSN)引入到工业算力网络中^[8,9]。当用户的业务需求从入口路由进入网络时,首先对业务流进行优先级划分及确定性需求识别,随后将业务封装为报文。在最高级别关键路由节点汇聚全网业务流的报文摘要,并按优先级分别进行排队,为各优先级的队列设置门控,结合时间流表和带有一定输出策略的输出选择器,将任务流转发控制信息在

网络进行广播，实现业务流的确定性实时传输。

4.3 算力解构、算力路由及资源调度技术

工业场景计算任务往往要求实时处理，对算力网络的时延和计算性能提出了挑战，现有的单个算力设备计算能力有限，难以缩短工业计算任务的处理时延。对此，将算力需求解构并基于算力路由技术调度多算力资源并行处理成为了未来工业计算任务的解决方案。

对于复杂的算力需求，如图3所示，可引入图聚类方法将需求解构为多个“子需求”^[10]，对算网中依赖与需求关系进行距离量化，并通过聚簇细分实现需求的分解。算力路由技术可为子需求寻找最优节点与网络转发路径^[11]，SRv6动态路由优化技术通过功能增强与扩展后，可满足动态多实例的算力路由需求。此外，基于Floyd算法的资源调度技术^[12]，通过对算力网络多源点加权图的分析完成资源调度，为实现协同多个算力终端节点并行处理任务提供了保障。

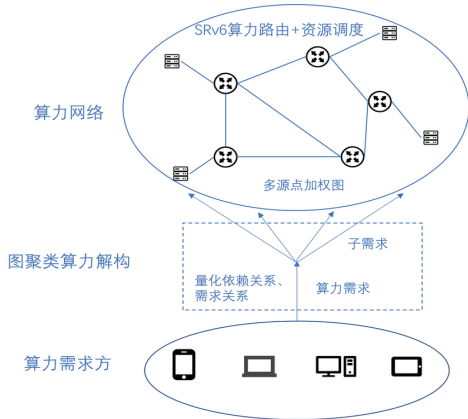


图3 算力解构、路由及资源调度

4.4 算力网络资源定价问题研究

由于工业算力网络的内外网资源分属不同的提供商，因此当工业智能任务需要算力资源的处理时，一个完备的算力网络资源交易和定价的体系是必不可缺的。

在工业场景下的算力网络中，智能计算任务往往对网络有着不同的时延需求。与边缘定价问题相比，通信、算力等因素相互影响的多资源定价问题亟需解决。网络中存在的通信干扰，会导致传输时延和功耗的增加；网路中的排队等待时间也会增加计算任务处理时延。对此，可考虑引入广泛用于资源定价的博弈论方法^[13,14]，联合算力需求与网络需求建模，构建服务提供

方与购买方的博弈模型并进行求解，同时实现智能设备的最大化效用与算力服务商的最大化收益。为了保证算力交易过程的安全性，可引入智能合约技术为交易过程中的数据提供隐私保护，最终实现可信环境下的算力交易。算力网络资源定价如图4所示。

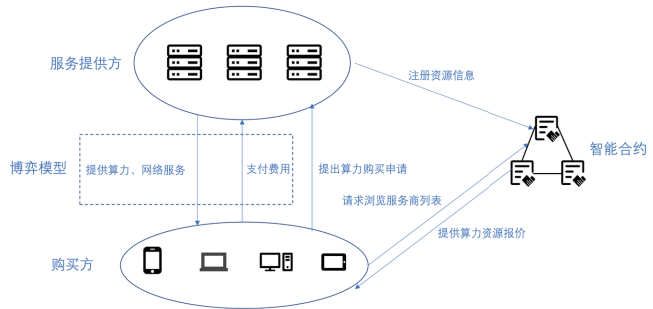


图4 算力网络资源定价

4.5 多终端协同的数据压缩技术

为了在算力网络中实现有效的数据传输，充分利用算力网络中各个节点的有效算力资源，基于信息论、编解码理论、学习理论有效地压缩海量环境数据，将是算力网络中的一项关键技术研究。

随着深度学习理论研究的快速发展，其在数据压缩领域的优势也逐渐凸显。随机神经网络、卷积神经网络、递归神经网络、生成对抗网络、变分自动编码器等陆续被应用到数据压缩技术之中，利用深度学习模型的非线性映射能力，可以快速学习到海量数据的深层次特征，更加有效地去除冗余数据特征，进而满足海量数据压缩的需求，实现算力网络节点之间海量数据的低时延传输。

4.6 数字孪生算力网络自优化技术

数字孪生技术是指利用数字化技术，为真实的物理实体创建对应的虚拟模型并模拟其行为，从而扩展与优化物理实体的功能。数字孪生技术目前已被引入网络的运维、管理、优化中^[15-17]，通过构建可实时交互映射的网络虚拟孪生体，最终实现对实体网络的优化管理。

在工业算力网络的场景下，可通过数字孪生技术与空间物理系统的结合，完成虚拟孪生体的创建。面对网络中节点超时的情况，算力网络提供超时节点与网络状态信息的实时采集，并调用孪生模块中的算法进行网络优化与验证，最终实现工业算力网络的自优化与运行管

理, 从而提升算力网络的运行效率并降低运维的人工成本。数字孪生算网自优化及管理如图5所示。

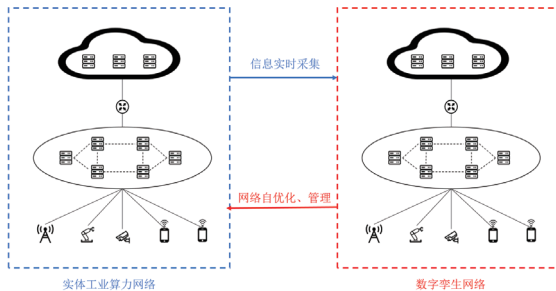


图5 数字孪生算网自优化及管理

5 结束语

工业互联网在向数字化、智能化方向发展的过程中, 对算力的高效利用提出了更高的要求, 这促使工业互联网与算力网络进行深度融合。本文在工业算力网络研究的初始阶段, 对其架构及运作思路、面临的挑战和实现所需的关键技术进行了总结和构想, 为相关研究人员提供了参考。**AP**

作者简介:

许士勋 (1999-), 男, 硕士, 现就读于北京邮电大学, 主要研究方向为算力网络、工业互联网。

张 奥 (1999-), 男, 硕士, 现就读于北京邮电大学, 主要研究方向为算力网络、工业互联网。

宋闻萱 (1998-), 女, 硕士, 现就读于北京邮电大学, 主要研究方向为算力网络、工业互联网。

黄潇洁 (1999-), 女, 硕士, 现就读于北京邮电大学, 主要研究方向为算力网络、工业互联网。

许方敏 (1982-), 男, 副教授, 博士, 现就职于北京邮电大学, 主要研究方向为物联网网络、未来网络技术。

赵成林 (1964-), 男, 教授, 博士, 现就职于北京邮电大学, 主要研究方向为短距无线传输技术、认知无线电技术、毫米波技术、工业互联网网络。

参考文献:

- [1] 边缘计算产业联盟, 工业互联网产业联盟, 网络5.0产业和技术创新联盟. 5G时代工业互联网边缘计算网络白皮书[R], 2021.
- [2] 唐静, 解云鹏, 雷波, 赵倩颖. 工业互联网场景下边缘计算的关键技术探讨[J]. 自动化博览, 2022, 39 (2): 34 - 37.
- [3] 中国移动通信集团终端有限公司, 北京邮电大学, 中国通信学会. 端侧算力网络白皮书[R], 2022.
- [4] 李建飞, 曹畅, 李奥, 等. 算力网络中面向业务体验的算力建模[J]. 中兴通讯技术, 2020, 26 (5): 6.
- [5] 郭凤仙, 孙耀华, 彭木根. 6G算力网络: 体系架构与关键技术[J/OL]. 无线电通信技术, 2023: 1-10. <http://kns.cnki.net/kcms/detail/13.1099.TN.20230104.1235.007.html>.
- [6] 周舸帆, 雷波. 算力网络中基于算力标识的算力服务需求匹配[J]. 数据与计算发展前沿, 2022, 4 (6): 20 - 28.
- [7] 刘鹏, 李志强, 陆璐. 面向工业互联网的算力网络思考[J]. 自动化博览, 2022, 39 (2): 25 - 28.
- [8] 贾庆民, 胡玉姣, 张华宇, 彭开来, 陈平平, 谢人超, 黄韬. 确定性算力网络研究[J]. 通信学报, 2022, 43 (10): 55 - 64.
- [9] 李宗辉, 杨思琪, 喻敬海, 邓仰东, 万海. 时间敏感网络中确定性传输技术综述[J]. 软件学报, 2022, 33 (11): 4334 - 4355.
- [10] 魏华, 张婷婷, 李莹. 算力网络一体化服务架构与实践[J]. 通信世界, 2022 (16): 39 - 42.
- [11] 黄光平, 史伟强, 谭斌. 基于SRv6的算力网络资源和服务编排调度[J]. 中兴通讯技术, 2021, 27 (03): 23 - 28.
- [12] 孙钰坤, 张兴, 雷波. 边缘算力网络中智能算力感知路由分配策略研究[J]. 无线电通信技术, 2022, 48 (01): 60 - 67.
- [13] 吴雨芯, 蔡婷, 张大斌. 移动边缘计算中基于Stackelberg博弈的算力交易与定价[J]. 计算机应用, 2020, 40 (09): 2683 - 2690.
- [14] 廖皓琳. 面向QoS按需的多用户Stackelberg博弈与网络资源定价研究[D]. 华南理工大学, 2021.
- [15] 王达, 孙滔, 孙晓文, 郭建超, 孙杰, 陈果, 杨爱东, 欧阳晔. 数字孪生在网络全生命周期管理中的研究[J]. 电信科学, 2022, 38 (04): 138 - 145.
- [16] 刘伟轩, 刘晓莲. 基于数字孪生技术的网络运维管理系统架构研究[J]. 信息与电脑 (理论版), 2021, 33 (07): 16 - 18.
- [17] 王奇奇, 莫皓颖, 户江民, 牟明铭. 基于数字孪生的网络优化方法研究[J]. 电声技术, 2021, 45 (01): 52 - 54.