



北京 2022 年 冬 奥 会 官 方 合 作 伙 伴
Official Partner of the Olympic Winter Games Beijing 2022

中国联通算力网络白皮书

中国联通网络技术研究院

2019 年 11 月

中国联通网络技术研究院

序

自电话发明并组网通信至今，已有百年，在这期间，运营商一直提供面向连接的基础网络，先后支撑了话音、专线、短信等电信级产品，业务与网络紧密耦合，运营商处于价值链主导位置，获得了良好的收益；然而近 10 年来，伴随着宽带和云的大发展，运营商却只提供了带宽（流量）服务，与个性化、非标准的互联网业务渐行渐远，被迅速管道化，由于管道不再紧贴用户需求，又被迅速低值化。从历史看，网络只有能够给应用赋能，才能与业务发展形成良性循环。

未来运营商面临从服务消费互联网到服务产业互联网的角色转变。5G 时代，运营商的网络资源和计算资源都将更加丰富，将成为全社会 ICT 基础设施提供方。同时，人工智能（AI）已经成为全社会的焦点，智能社会正在快速到来。在 5G+AI 的产业背景下，是否在带宽（流量）业务之外，还存在适合运营商经营，标准归一化的新业务形态？

“算力网络”将尝试回答这个问题。

中国联通网络技术研究院

目 录

1	产业背景	1
1.1	机器智能社会将全面到来	1
1.2	网络将出现云、边、端三级算力架构	2
1.3	实现云、边、端算力的高效需要算力网络	2
1.4	运营商的可持续发展需要算力网络	4
2	算力网络的概念和架构	7
2.1	算力网络是云化网络发展演进的下一个阶段	7
2.2	算力网络的关键技术元素	8
2.2.1	联网元素：打造无损和确定性的网络联接	9
2.2.2	云网元素：智能网络与网络云化的持续推进	10
2.2.3	算网元素：为计算服务的可信、高效、随需网络	10
2.3	算力网络的典型应用场景	16
2.3.1	运营商 ToB 的“5G 园区+AI”场景	16
2.3.2	运营商 ToC 的“5G+Cloud X”场景	17
2.3.3	算力开放，运营商提供可交易的算力通证	18
3	算力网络的标准与生态	19
4	总结与展望	20
5	缩略语	22

中国联通网络技术研究院

版权声明

本白皮书版权由中国联通网络技术与华为技术有限公司共同拥有，并受法律保护。转载、摘编或利用其它方式使用本白皮书文字或者观点的，应注明来源。

作者

中国联通网络技术研究

唐雄燕、曹畅、张帅、何涛、刘莹

华为技术有限公司

杜伟、郑若滨、唐晓琴

中国联通网络技术研究院

1 产业背景

1.1 机器智能社会将全面到来

人类将步入智能社会，智能是知识和智力的总和，翻译到数字世界就是“数据+算力+算法”，其中算法需要通过科学家研究实现，海量数据来自于各行各业的人和物，数据的处理需要大量算力，算力是智能的基础平台，由大量计算设备组成。

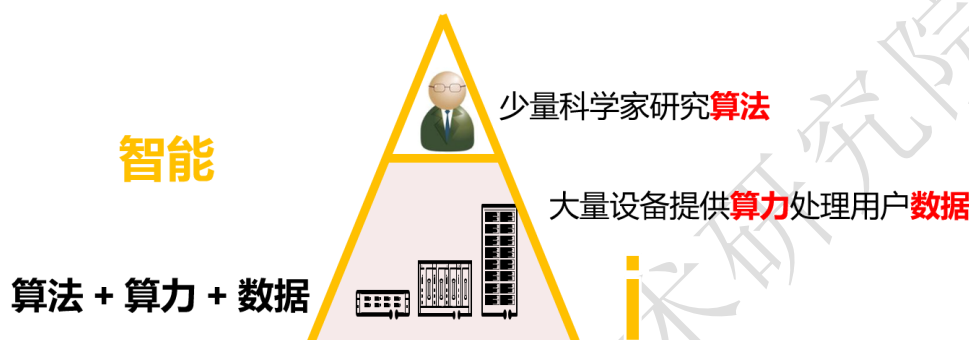


图 1-1 智能的三要素：算法、算力、数据

人脑的算力相当于约 300 亿颗晶体管，人类的历史和文明都是由无数人脑算力所创造，但人脑算力正面临老龄化的挑战，2020 年超高龄国家（65 岁以上人口超过总人口 20%）将达到 13 个，2030 年将上升到 34 个，而且主要集中在亚太、欧美等较发达国家。考虑到儿童占比约 15%，实际这些国家的劳动适龄人口只占不到 60%。全球 2020 年人口约 77 亿，较发达国家人口约 30 亿，所以这些国家实际处于劳动适龄段的人脑只有不到 20 亿。

现阶段电子工艺可以做到的机器算力已经接近人脑算力，如麒麟 980 基于 7nm 工艺集成了 69 亿晶体管，AMD Radeon VII GPU 将采用 7nm 工艺，晶体管数量约 132 亿，未来 5 年，基于 5nm 工艺，芯片集成度据信可以做到 300 亿晶体管，此时处理器的信息分析处理能力已经与人脑相当，并且相比于人脑，处理器更聚焦于专业领域的数据处理，不知疲劳，所以在具体的数据处理领域，高端 CPU 的算力已经事实上相当于甚至于超过了人脑。

机器替代人处理数据是人类进入智能社会的必然趋势，目前全球应用处理器出货量已经超过 20 亿，并且仍以 8% 的速率在高速增长，面临老龄化的国家（亚太和欧美国家）同时也是半导体主要的消费市场，占全球份额约 80%，5~10 年之后，这些国家的机器算力就将全面超越人脑算力的总和。

2020 年人脑为主，机脑为辅，以各种智能终端为代表；2025~2030 年人脑

和机脑相互协助,智能机械臂,辅助驾驶,智能制造等广泛出现;2030~2040年,将逐渐进入“人脑定义规则,机脑自主执行”的“机器智能”社会,机器管家、无人驾驶、无人工厂等广泛出现。

1.2 网络将出现云、边、端三级算力架构

全球数据总量仍在持续增长,预计2020年达到47 ZB,2025年达到163 ZB,年复合增长率为20%,其中绝大多数来自亚太(约40%)和美(约25%)、欧(约15%)地区。全球数据中心安装服务器数量2020年将达到6200万台,年增长约4%;智能终端(包含手机/M2M/PC等)年复合增长约10%;由于工艺的约束,单芯片的算力在5nm之后将接近顶峰,传统集约化的数据中心算力和智能终端的算力可增长的空间也面临极大挑战。要支持数据持续增长的机器智能时代,只有终端+数据中心两级处理无法满足要求,算力必然会从云和端向网络边缘进行扩散。数据处理会出现三级架构:终端、边缘和数据中心,边缘处理能力未来几年将高速增长,尤其是随着5G网络的全面建设,其大带宽和低时延的特征,将加速算力需求从端、云向边缘的扩散。

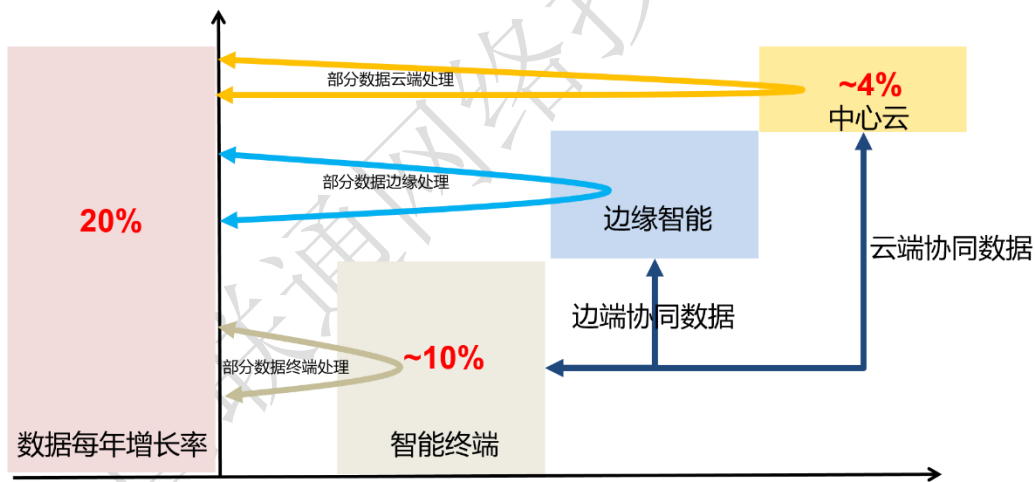


图 1-2 云、边、端三级算力架构

1.3 实现云、边、端算力的高效需要算力网络

高效的定义是“在相同或更短的时间里完成比其他人更多的任务,而且质量与其他人一样或者更好”。一个终端或者在一个数据中心内部做到高效算力相对容易,但对于广泛分布在云、边、端的算力如何做好协同,并实现高效则充满挑战。

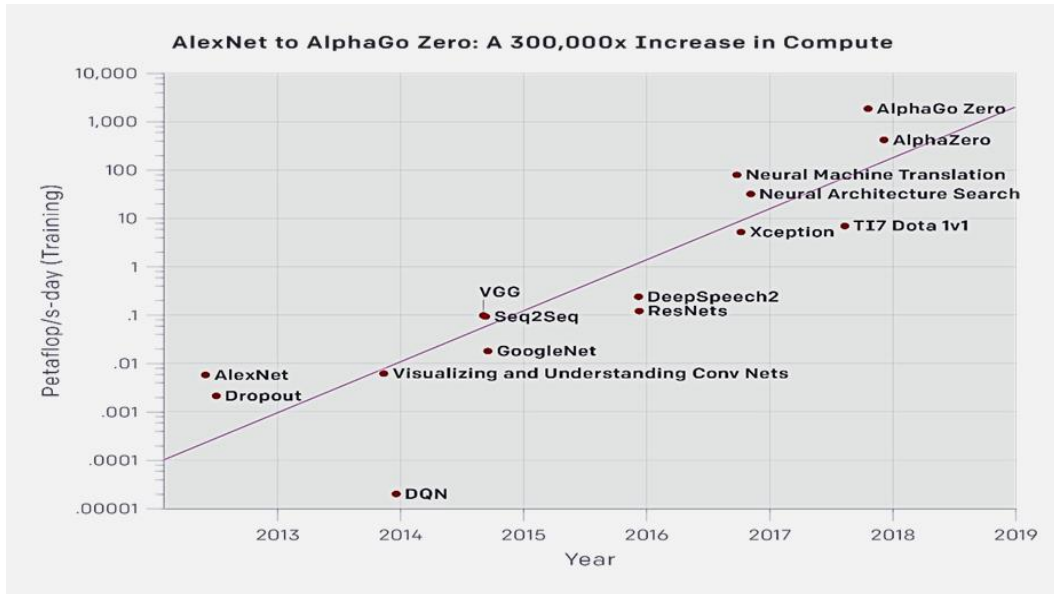


图 1-3 AI 训练中不同算力对算法的需求

算力按照应用场景有不同的衡量单位，用于比特币的每秒哈希运算次数 (H/S)，用于 AI 和图形处理的每秒浮点运算次数 (FLOP/S)，智能社会对算力的诉求主要是浮点运算能力，专用 AI 芯片如华为昇腾 910 采用 7nm 工艺，半精度 FP16 算力达 256 TFLOPS，低功耗的 12nm 芯片昇腾 310 半精度 FP16 算力也达到了 8 TFLOPS。过去 5 年，随着深度学习算法的演进，AI 训练对算力的需求增加了 30 万倍，一些互联网厂家已经将算力作为服务提供给客户，从 1 FP32 TFLOPS 或 8 FP16 TFLOPS 到 4 FP32 TFLOPS 或 32 FP16 TFLOPS 的 AI 推理加速服务，简单的语音语义识别或单流视频分析 8 FP16 TFLOPS 即可满足，复杂的推荐引擎或者风险检测则需要 32 FP16 TFLOPS。

表 1-1 AI 推理不同应用对算力的需求

AI 推理场景	算力分类	32 位浮点运算	16 位浮点运算	内存
语音语义或者单流视频	中等	1 TELOPS	8 TELOPS	1 GB
多流视频	大型	2 TELOPS	16 TELOPS	2 GB
推荐引擎或者风险预测	超大型	4 TELOPS	32 TELOPS	4 GB

高效算力的第一个要素是“专业”，聚焦专用场景可以用更低的功耗和成本完成更多的计算量。目前相对成熟，而且对算力需求最大的场景是视频和图像分析

领域，基本上贯穿了所有行业的智能场景，在边缘和云进行视频的分析 and 处理，都需要网络提供高吞吐能力，网络吞吐能力取决于网络带宽和时延两个关键指标，带宽越大，时延越低，数据吞吐量越大。

高效算力的第二个要素是“弹性”，如算力的超分配、算力提供的敏捷性，如轻计算 100ms 级弹性，数据弹性处理需要网络为数据需求到算力资源之间提供敏捷的连接建立和调整能力。

每个人的高效工作不一定就能构成一个高效社会，**高效算力的第三个要素是“协作”**。从处理器内部多个核之间的协作，到数据中心内部多台服务器之间的“算力均衡”，再到整个网络边缘的“随选算力”。协作的目的是实现算力资源充分被使用，需要承载网络支持多边缘之间、边缘与中心之间的算力均衡，流量调度和拥塞管理。

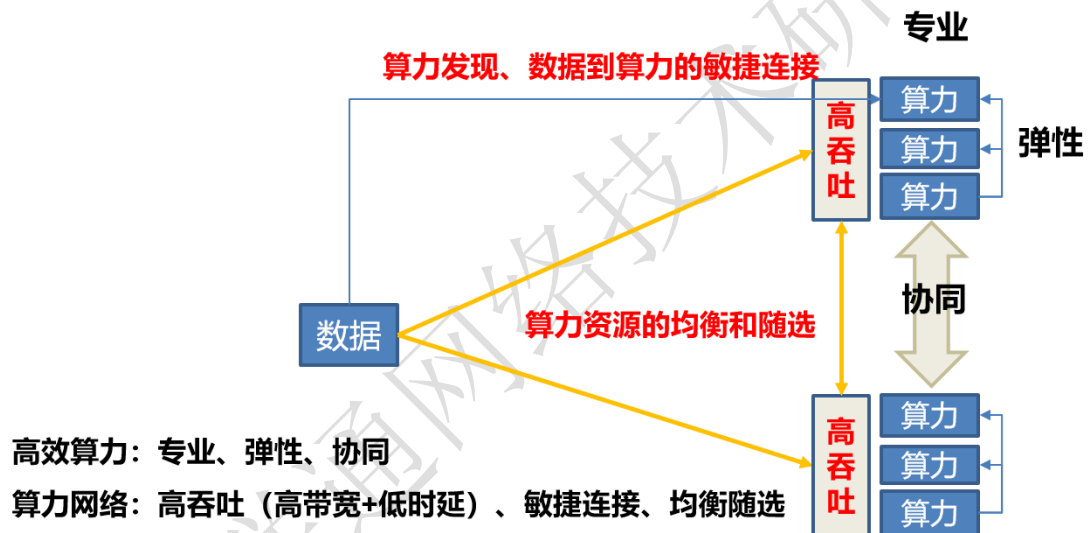


图 1-4 高效算力的三个要素

实现高效算力需要“计算+网络”深度融合的新型网络架构“算力网络”，实现数据与算力的高吞吐、敏捷连接和均衡随选。

1.4 运营商的可持续发展需要算力网络

运营商网络的核心价值从近 20 年运营商的发展来看，一直是业务和技术双轮驱动，且以移动通信的代际划分为标志，从 2G 到 4G，运营商提供的业务和为提供这些业务所采用的支撑技术，均在不断发生变化，而这些变化也直接影响着运营商的经营状况和收入水平，体现出运营商在不同时期的价值。

从业务驱动来看，面向最终个体消费者的 ToC 业务，特别是移动业务一直

是运营商最大的收入来源。在 2G 时代，运营商的 ToC 业务体现为语音、短信等。这一类业务的特点是：用户需求刚性，与网络连接能力强相关，业务即提供最终服务，且从连接到业务的全部环节有运营商完全掌控。因此，在 2G 时代，全球移动运营商均获得了丰厚的收益。但是，当经过 3G 时代的短暂过渡，历史最终跨入 4G 时代后，由于连接技术的飞速发展，网络带宽支撑一般的个人应用已完全没有问题。同时，随着云计算、智能终端的崛起，运营商 ToC 端主要经营的业务由语音、短信变为了流量，虽然流量的提供也需要运营商网络的良好保障，但是，由于流量不是最终的业务形态，而只是 OTT 提供业务的带宽资源，导致运营商与客户的最终业务，最直接需求越走越远，加之自身研发能力和管理水平的限制，导致运营商在云时代迅速被管道化，价值越来越低。依靠流量经营，增量不增收，是目前全球运营商所面临的共性问题 and 严峻挑战。

从技术驱动来看，运营商一直是全球 ICT 行业技术的牵引者，是电子信息行业全产业链的重要参与者。最先进的 ICT 技术，往往需要运营商为其快速提供网络支撑，从而形成商业能力，而这些先进的 ICT 技术，反过来又促进电信网络的不断革新和进步，增强了运营商提供服务的能力。例如，在已经到来的云计算时代，运营商网络作为云内部和云间的“总线”，为云服务的正常运行提供了强有力的支撑，同时，自身网络也在以 SDN/NFV 为代表的网络云化的影响下发生着深刻的变革，网络的虚拟化进程不断加快，跨网络专业的调度和编排能力显著提升，未来 5G 的核心网网元将全部承载在通信云上，网络实现了真正的云化。

展望未来，在大数据、物联网、云计算等新兴技术的影响下，全社会的信息化进程明显加快，企业的数字化转型愈加提速，ICT 赋能的社会表现出以下几个趋势：

- **ICT 业务的主要增长点由消费互联网转移到服务产业互联网**，运营商除了给个人用户提供服务，给政企客户提供专线服务外，更要向全产业赋能。
- **电信网云化后，在满足自身云服务之外，有希望对外输出计算能力，成为全社会的 ICT 基础设施**。5G 时代，随着运营商网络的云化改造，将有大量的 IT 资源应用在通信云的部署中，而且随业务量变化会存在一定时间段内的空闲，运营商可以将这些资源，结合其强大的网络提供服务能力对外服务，并且比公有云服务更加靠近用户。
- **AI 技术成为全社会的焦点**，在各行业 AI 应用的加持下，一个崭新的 AI

社会将快速到来。未来面向全社会的 AI 普适服务，离不开 AI 算力的支持。从运营商以往的经验来看，运营商并不完全适合经营碎片化、一企一策的定制化服务业务，而是更适合经营标准化、归一化、规模化，具有网络强连接属性的业务，并通过这样的业务为全社会的数字化转型赋能。

结合个人到企业，全社会对 AI 服务的迫切需求，以及运营商强大的网络服务能力和空闲的 IT 资源，完全可以向全社会提供普适性的 AI 算力，使算力成为继话音、短信、专线、流量之后运营商提供的新一代普适性标准化产品。且该产品的价值比流量更高，可以结合用户的时延需求，计算能力与计算位置需求，充分结合运营商 5G 网络接入优势，边缘机房和光纤覆盖优势，以及富余 IT 资源优势，包装成面向不同用户层级、不同需求的多等级、多种类算力服务，为 AI 社会提供强大的“能源”与“养料”。

结合上述分析，运营商在算力富余、全社会 AI 需求旺盛的背景下，更适合为广大政企客户提供赋能 AI 应用的算力服务，因此，运营商下一代网络也必须按照提供算力服务的要求建设，打造新时代的**算力网络**。

2 算力网络的概念和架构

网络的核心价值是提高效率，电话网提高了人类沟通的效率，互联网提高了人类协同工作的效率，算力网络的出现是为了提高端、边、云三级计算的协同工作效率。区别是传统网络直接为人类服务，算力网络直接为智能机器服务，并通过智能机器间接为人类服务，由于机器的工作效率、响应时间都远超人类，因此对算力网络会有更严苛的要求。

2.1 算力网络是云化网络发展演进的下一个阶段

云化网络的概念是从 2009 年开始，伴随着 SDN/NFV 技术而出现，并不断发展和演变，早期云网协同阶段网络主要是为云计算提供联接服务（Network for Cloud），包括了基于 SDN 的 DCN 和 DCI 网络和用于用户与云联接的 SD-WAN Overlay 网络；同时网络也在运维上借鉴了云计算的技术（Cloud for Network），在业务发放效率上做出了提升。

2014 年开始，伴随着电信行业对云技术理解的加深，开始思考如何利用云的技术理念来改造网络设备，网络云化进入云网一体阶段，其中包括了对核心网网元进行云化改造（NFV），对承载网架构进行转控分离的改造，并实现转发面的极简（如 SRv6）和小型化（如 C/U 分离的核心网）的成功实践，经过 10 年的探索，已经基本上确定了“控制云化+转发极简”的云化网络标准架构。在这个过程中全球运营商也基于各自网络特点进行了探索，这其中包含了 AT&T Domain2.0、中国联通 CUBE-Net2.0、中国电信 CT2025、中国移动 NovoNet 等成功实践。

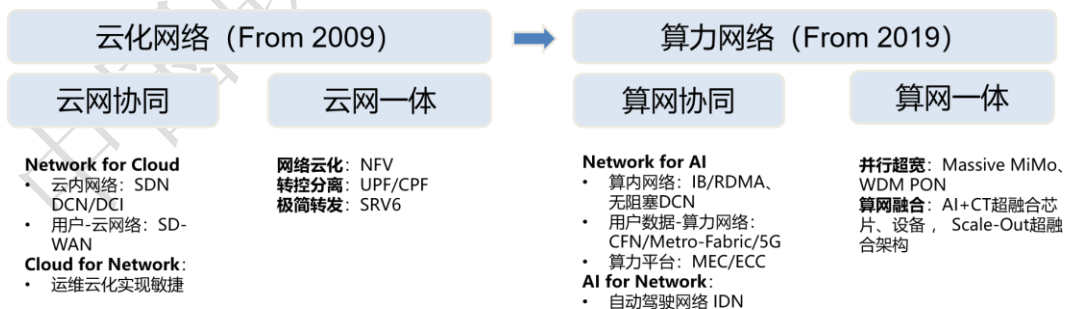


图 2-1 云化网络和算力网络技术内容对比

2019 年，伴随着 AI 的快速发展，高效算力成为支撑智能化社会发展的关键要素，并开始在各行各业渗透，算力网络作为一种新型网络架构逐渐被业界认可，并有可能成为网络演进的新方向。

算力网络早期与云化网络类似，聚焦算网协同的需求，包括网络为 AI 提供联接服务（Network for AI），如用于数据中心内部的算内网络 IB、RDMA 和更大规模的无阻塞 DCN；用于用户数据到算力联接的 CFN（Compute First Network）和支撑用户数据到算力更低时延和更大带宽的 Metro Fabric 和 5G URLLC 网络；用于为 AI 提供算力服务的新型网络设备 MEC/ECC 等；还包括将 AI 技术用于运维（AI for Network）实现主动运维的 IDN 自动驾驶网络。

伴随着网络对 AI 技术的深入研究，未来 AI 的一些核心理念如“并行低精度+算法算力辅助”有可能会用于网络设备、芯片的改造，突破单通道物理极限的挑战，通过多通道+算法算力的方式实现更大接入带宽，因此一些新型超融合（CT+AI）设备可能会在网络中大量出现，算网协同也将向算网一体阶段演进。

算力网络中算网协同的场景已经逐渐出现，与云网协同类似，首先面向数据中心内的需求，然后向用户到数据中心跨广域的需求延伸。算网一体多数还处于研究阶段，未来 3~5 年会逐渐成熟。

2.2 算力网络的关键技术元素

高效算力网络，具备了联网、云网与算网三个方面的技术元素，其中联网是基础，在 5G 时代，引入了超低延时技术与端到端网络确定性技术，以适应 VR/AR，工业计算等面向垂直行业的需求。同时，为了达到网络无损的目的，需要将数据中心内部的 Leaf-Spine 架构向城域扩展，搭建城域的 Metro Fabric。云网方面，网络人工智能技术将在算力网络的运维、管理、故障预测等方面发挥极大作用，并且网络需要进一步的云化以便提升业务交付效率。算网技术元素，主要包括算力生成、算力调度（路由）和算力交易三个方面，使网络成为为全社会提供 AI 算力能力的基础设施。

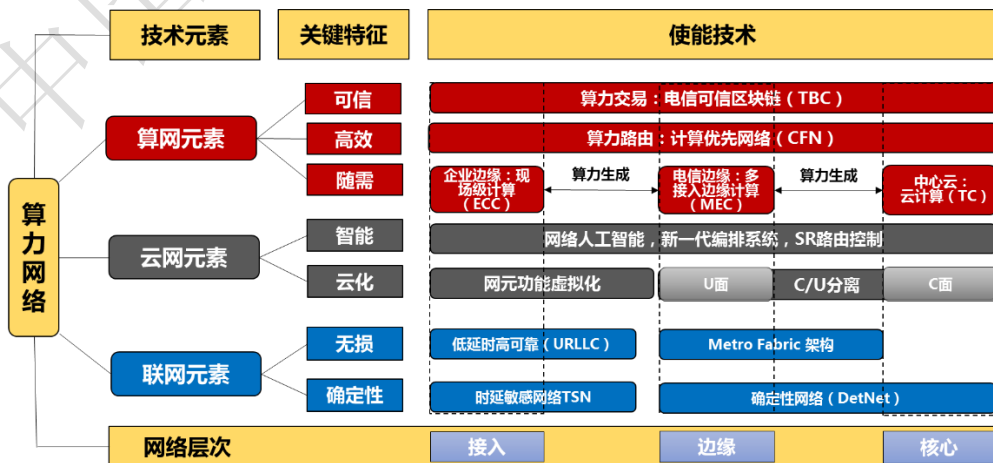


图 2-2 算力网络的关键技术元素

2.2.1 联网元素：打造无损和确定性的网络联接

低时延高可靠（URLLC）：除传输距离带来的光纤时延之外，设备层面对时延影响最大的是无线接入网，数据处理时延在 ms~10ms 级；其次是数据通信和云核心网络，轻载时延在 1ms 量级，但是如果出现拥塞的话，时延会急剧攀升数十倍；影响最小的是光传输，时延在 100us 级。因此在时延问题上，主要矛盾在于无线接入如何降低时延、承载网与核心网如何避免拥塞，提供确定性时延，URLLC（Ultra-Reliable and Low-Latency Communication）超级可靠低时延联接是一个端到端的概念，这里仅指 5G RAN 范围的 URLLC，在 3GPP RAN TR38.913 中定义了 URLLC 的指标，控制面板时延 10ms、用户面时延 0.5ms、移动性中断时间 0ms、可靠性 99.999%（32bytes@1ms）。

无阻塞城域网（Metro Fabric）：数据通信网络解决低时延的主要方案是避免拥塞，从“话音/宽带业务为中心”到“云/计算业务为中心”，老的转发技术和复杂网络已经不适应这种变化，需要进行重构，云计算对 DC 的要求从“一个 DC as one Computer”已经发展为“多个 DC（中心和边缘）as one Computer”，相对应的承载网也需要从“DCN Fabric 是一台无阻塞路由器”到未来“整个城域网 Metro Fabric 是一台无阻塞路由器”演进。

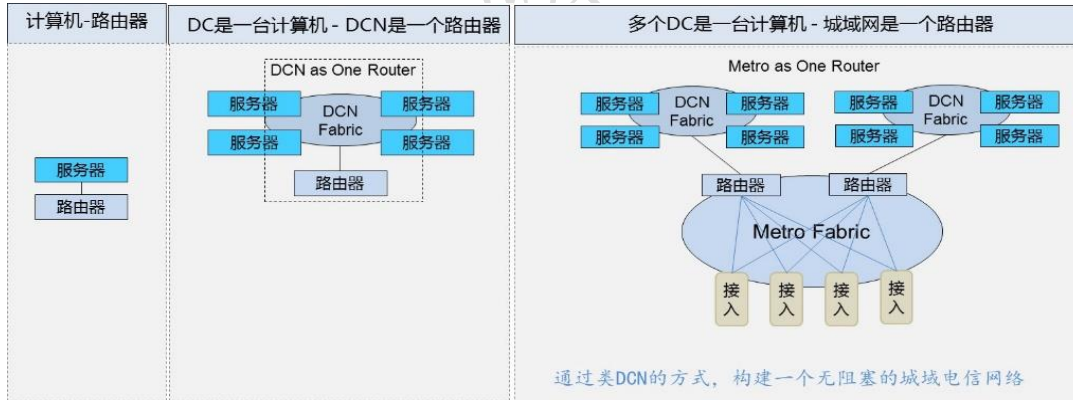


图 2-3 无阻塞城域网 Metro Fabric 架构

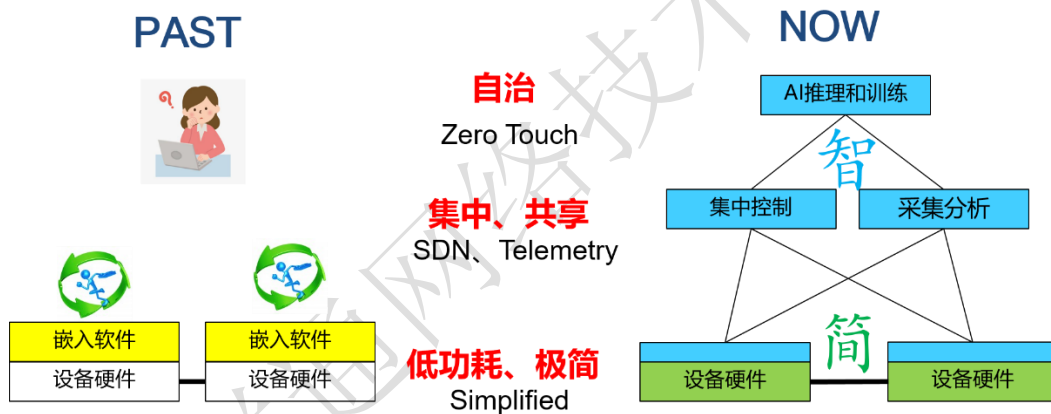
为支持 Metro Fabric，首先需要改变城域网多级汇聚收敛的架构，构建用户接入到云（中心或边缘）一跳直达的扁平化网络，通过减少跳数，实现网络拥塞点的减少。在 Fabric 架构下，网络拥塞主要矛盾出现在存在多方向流量的业务边缘节点，这个位置可以采用支持无阻塞交换算法的超大容量路由器实现。城域网向 Metro-Fabric 架构演进，实现小于 1ms 的确定性网络延时的能力。

确定性网络：DetNet（确定性网络）指在一个网络域内给承载的业务提供确定性业务保证的能力，这些确定性业务保证能力包括时延，时延抖动，丢包

率等指标。DetNet 可以通过多个路径发送序列数据流的多个副本，并消除目的地或附近的副本。通过调整数据包的传送并为临界流（Critical Flow）分配足够的缓冲区空间，可以消除拥塞。不存在故障检测和恢复周期，每个数据包都被复制并被带到或接近其目的地，因此单个随机事件或单个设备故障不会导致丢失任何一个数据包。

2.2.2 云网元素：智能网络与网络云化的持续推进

控制面智能：随着转控分离、控制功能云化和三级分布式部署带来网络灵活性的同时，也带来了集成和运维复杂度的快速攀升，传统人工运维的方式已经难于支撑，未来将逐渐引入大数据分析和人工智能来替代人工（AI for Network），但是过度复杂的网络协议也会增加 AI 算法实现的难度，除了控制面智能化外，还需要对转发面的协议和网络拓扑进行简化，智简是“网络自动驾驶”的基础。



智：AI替代人，维护复杂的网络关系；简：简化网络关系，降低运维复杂度

图 2-4 人工智能技术在网络中的应用

网络云化：网络控制面包含实时功能（如可靠性，算法，协议处理相关）和弱实时功能（如管理、配置相关），弱实时功能最先云化，可以灵活的部署在云端，而实时功能云化后，需要集中控制的功能可以靠近被控制设备，部署在网络中心或者边缘云，而一些实时分布式控制功能，还将进一步以微服务形式嵌入在设备内部。层次化的云化控制功能既保留了云化带来的敏捷和开放性，也可以按需满足实时性的系统要求。

2.2.3 算网元素：为计算服务的可信、高效、按需网络

业务层网络由业务处理节点和节点之间的连接组成，未来网络要从信息传

输为核心的信息基础设施，向融合感知、传输、存储、计算、处理为一体的智能化信息基础设施发生转变(来源：中国人工智能 2030 战略规划)，这对业务处理节点的功能、节点之间的连接技术都提出了新的要求：业务节点从只处理电信业务的封闭模式向可对外开放，提供开放算力服务能力的新形态节点发展，典型的如边缘计算 MEC (Multi-access Edge Computing)，连接从“对业务无感知，私有网络”向“感知用户业务需求，为数据和算力服务之间建立按需连接”的开放型网络发展，典型的如计算优先网络 CFN (Computing First Network)；用户的管理认证、网络资源开放与服务化需要从传统中心化、静态方式向去中心化、动态、安全、可信新方式发展，典型的如电信可信区块链 TBC (Trusted Block Chain)。

随需的边缘计算：MEC(Multi-access Edge Computing)即多接入边缘计算，其中的“M”- Multi-access 包含 MBB、WLAN 等多种接入方式，可以看作是一个运行在电信网络边缘的、运行特定任务的云服务器。MEC 为应用开发者和内容提供商提供了超低时延、高带宽，以及可以实时获取无线网络信息的业务环境，从而为终端用户提供差异化的业务和服务。

MEC 部署在靠近基站的接入环、接入汇聚环等边缘位置，使得内容源最大程度的靠近终端用户，甚至可以使终端能够在本地直接访问到内容源，从数据传输路径上降低了端到端业务响应时延。例如，AR/VR 业务要求端到端时延需小于 20ms，以消除用户的眩晕感，5G 更是提出了 1ms 端到端时延来支撑自动驾驶等时延敏感业务。MEC 通过将对应的网络功能部署在最靠近用户的边缘位置，使业务达到极致的体验。据研究，未来有 70% 的互联网内容可以在靠近用户的城域范围内终结；基于 MEC，可以将这些内容存储在本地，MEC 与终端用户之间的传输距离缩短，流量在本地被卸载，节省了 MEC 到核心网和 Internet 的传输资源，进而为运营商节省 70% 的网络建设投资；目前，越来越多的细分领域希望基于电信网络实现行业定制，通过 MEC 提供开放的平台可实现电信行业和垂直行业的合作业务创新。ETSI 定义的 MEC 是具备无线网络能力开放和运营能力开放的平台，MEC 可通过公开 API 的方式为运行在其平台主机上的第三方应用提供无线网络信息、位置信息、业务使能控制等多种服务。

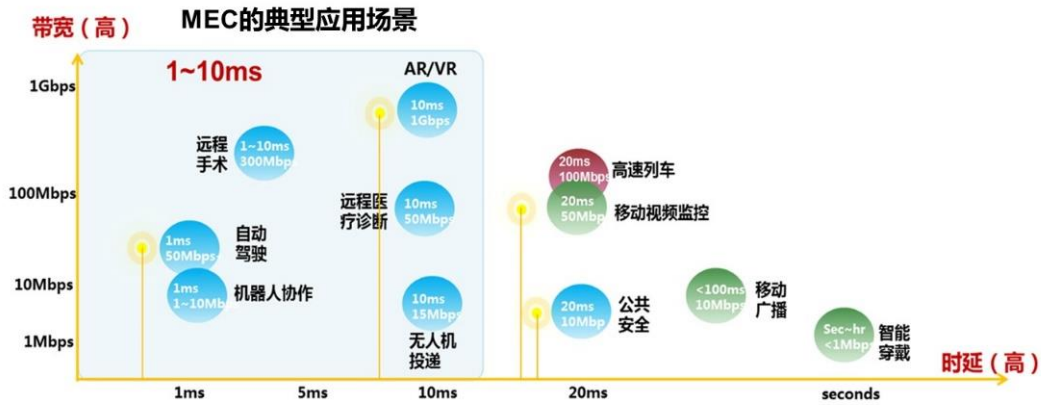


图 2-5 边缘计算典型应用场景说明

高效的计算优先网络：流量是互联网行业的核心竞争力。传统的互联网流量获取方式包括搜索引擎、聚合网站、移动分发、门户网站等等，比如日常接触到的百度、Google、新浪、安卓市场、苹果商店。同时流量也是运营商当前的主要收入来源。当前，在成本驱动下，如何低成本高效率地获取流量始终是互联网企业的一大难题，同时由于传统运营商网络无法感知互联网业务，无法匹配互联网业务需求多样性，因而无法提供差异化服务，大量可以按照业务属性进行优化，只需要本地处理和终结的流量，却配置为穿越整个骨干网到异地处理。互联网上 50% 的流量属于这种低效率的无序流量，大幅增加了运营商骨干网络扩容压力，提高流量效率对运营商和互联网企业都具有重要价值。

另一方面，计算技术正在向着轻量化、动态化、应用解构成服务/功能的方向变化和发展。计算载体从虚拟机，发展到容器以及 Unikernel，镜像大小由 VM 的 GB 级别，到容器的 MB 级别逐步降低到 Unikernel 的 KB 级别，实例化时间由分钟级到秒级再到 100ms 级，变得更加轻量，启动更快，运行代价更小。此外随着微服务的发展，传统的 Client-server 模式被解构，Server 侧的应用解构成功能组件布放在云平台上，由 API Gateway 统一调度，可以做到按需动态实例化，服务器中的业务逻辑转移到 Client 侧，Client 只需要关心计算功能本身，而无需关心 Server、虚拟机、容器等计算资源，从而实现 Function As a Service。未来应用无需感知 Server，计算生命周期缩短、地点动态变化的趋势使得互联网架构的业务假设发生变化。

其次，计算资源融入网络使得架构的拓扑假设也要发生变化。传统互联网架构的基本拓扑抽象是端到端模型：网络在中间、计算在外围，主机通过网络实现逻辑虚拟的全连接。而边缘计算或者泛在计算的场景中，拓扑变成了计算嵌在网络中间，从完成用户的计算任务的角度，嵌入的资源不再是对等的 Peering 关系，而是需要考虑距离的不同，以及网络状况的好坏。

边缘计算、乃至泛在计算场景中，由于单个站点的算力资源有限，需要多站点协作，现有架构一般通过集中式编排层来管理和调度，存在可扩展和调度性能差的问题。现有业务应用层和网络解耦，应用层无法精准、实时掌握网络性能，以应用层为主的寻址结果的综合性能可能不是最优，甚至比较差，导致业务体验差。此外，当前互联网的假设是静态的 Server 加上移动的 Client，传统基于 DNS 解析的 IP 寻址，以及建立 TCP/TLS 会话的网络模式，也难以发挥动态、微服务、泛在计算的优势，不能保证计算效率最大化。

未来网络架构，需要能够支持不同的计算功能，根据不同的业务需求，网络状况，可以在离 Client 的不同距离实时地实例化。基于分布式网络的计算网络融合新架构——CFN 试图解决以上问题。

计算优先网络(CFN)位于网络层之上的 CFN 薄层，将当前的计算能力状况和网络状况作为路由信息发布到网络，网络将计算任务报文路由到相应的计算节点，实现用户体验最优、计算资源利用率最优、网络效率最优。

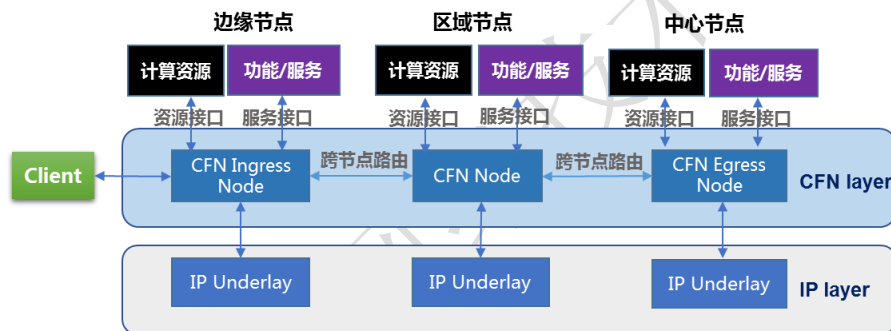


图 2-6 计算优先网络承载算力路由

CFN 具有以下关键技术特征：

- 构建计算与网络融合的新架构：云网融合、网随云动、云在网中，将用户尽快获得计算结果作为第一设计原则的全新网络架构。
- 计算路由器(CFN Router)连接分布式的计算集群联邦，其基本功能是基于请求中 Service ID 和 Data ID 的路由，高级功能还包括数据预取等能够提高效率 and 用户体验的功能。
- 定义 CFN 与服务间的通用计算接口，与应用无关，语义不感知，使得 CFN 与千变万化的应用创新解耦，保持自己的稳定。
- 开放路由协议：基于抽象后的计算资源发现，计算资源层面的拓扑和路由生成和愈合，拓扑和路由不仅仅是 IP 可达性，还包含计算资源的动态变化和可服务性，愈合不仅仅是路由的愈合，而是对一个计算任务的重新路由和调度，

使之能够在一个新的计算服务节点上完成计算。

CFN 能够为应用提供的核心价值是基于当前网络可用的算力和算法，结合网络实时状况，通过 CFN 灵活匹配、动态调度计算资源，将终端的计算卸载到合适计算节点、边缘或者中心云，支撑业务的计算需求，保证业务的用户体验。典型应用如 AR 场景，终端通过摄像头采集图像，在家庭网关和边缘节点进行物体识别中等复杂度的计算、以及需要尽快给出结果的计算，将 AR 场景中计算负载最重的 Object Recognition，次重的 Feature Extraction、Object Detection、Template Matching 等计算任务动态卸载到边缘或者中心节点，终端只需负责目标跟踪和画面显示，降低终端的计算开销。物联网（IoT）场景中物联网终端设备因资源受限，具有计算能力较弱、存储空间较小的特点，将终端侧采集的数据，在边缘节点进行数据预处理，减少上行传输的数据量，在中心云实现数据存储和数据同步等任务。利用边缘和中心节点的计算能力最大程度降低 IoT 受限终端设备的计算负载。

可信的电信区块链：区块链作为一项颠覆性技术，正在引领全球新一轮技术变革和产业变革，推动“信息互联网”向“价值互联网”变迁。区块链概念最早出现在中本聪的比特币白皮书中，但并非以“区块链”的提法出现，而是以“工作量证明链”的形式陈述。2008 年 11 月 1 日，中本聪发表《比特币：一种点对点的电子现金系统》一文，阐述了基于 P2P 网络技术、加密技术、时间戳技术、区块链技术等电子现金系统的构架理念。为了在通信领域应用该技术，ITU 相关标准中引入了“通证 Token”的概念，用于数据或资源的共享和交易。对于算力网络而言，为了方便算力的量化、共享和交易，将使用“算力通证”的概念。

区块链已经不仅仅是一项技术、一种工具，更是一种思维方式，区块链作为一种新型的技术组合，其去中心化、难以篡改、不可抵赖等特点不仅为电信行业带来了一种全新的信用模式，也使其数字服务更具竞争力，进而帮助电信行业降低成本，为该领域带来了全新的视角。

随着通信网络技术的发展，很多业务对宽带、时延等都提出了新的要求。为了提升用户体验，未来很多海量的数据都需进行本地化或网络边缘处理，这样可以降低网络负荷，并能获得更低的时延。运营商开放边缘计算的算力，在交易模式上可以考虑与区块链的结合，技术上采用联盟链将更加高效，也可更好地满足监管、审计的需要。在实际部署中，区块链平台或应用可以安装部署在边缘计算服务器上，为不同的应用场景提供区块链技术和能力支撑。

在移动网络边缘的 MEC，因为机房、环境等条件限制，MEC 的硬件资源往往比较受限，而在 MEC 周边有一些设备具有较强的处理能力，如手机/摄像头/个人电脑等，这些资源可能被利用起来强化 MEC 的能力。MEC 连接的本地网络服务器，或是本地网络中的一些个人电脑，均有一定的存储、算力资源。利用区块链技术，可以汇聚闲置的各类资源构建“无限节点”的资源网络，从而聚合成一个强大的资源池，并对分布在各节点的资源进行最优化的实时部署利用，这样可以有效地帮助 MEC 部署方节约成本和提高效率。MEC 与区块链技术的结合，可以提供丰富的算力等共享资源，用于视频直播、本地缓存等业务，GPU 资源也可以用于 AI 训练等。

MEC 的运营者可以采用链上积分或链下支付的方式进行回报。交易后 MEC 可以操作使用接入区块链上的相应的资源。下图给出了一种设想的 MEC+区块链的资源交易过程。

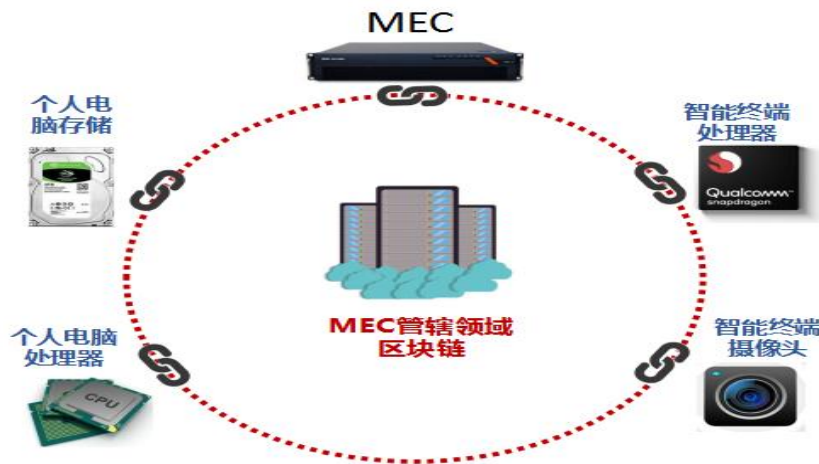


图 2-7 MEC+区块链资源交易过程示意

- 移动用户在设备上安装所需软件，将设备信息上报到 MEC，包括设备类型、资源配置、使用情况等，由 MEC 将其记录在链上。固定设备的信息发送到链上进行登记。
- MEC 根据自身资源情况，或由运营者操作，使用其它设备的共享资源进行业务处理，将具体资源使用情况的数据放到链上，并给予共享设备者一定的积分奖励，记录到链上。
- 用户在购买 MEC 区块链范围内业务/产品时，可选择使用积分兑换，积分使用情况记录在区块链上。

基于区块链的特点，该场景实现了弱中心化的微服务和异步任务执行。同时，可以更进一步设想新的商业模式：计算资源需求方可把希望执行的任务发布出来，

任务被分割成碎片派发给 MEC+区块链链上的计算资源贡献者，贡献者执行碎片任务，获取相应奖励报酬，形成闭环。借助智能合约，MEC+区块链在资源提供方和资源需求方之间架起了桥梁。(来源:可信区块链推进计划电信行业应用组)

2.3 算力网络的典型应用场景

算力网络与 5G 和 AI 的技术成熟度存在密切关系，最先成熟的应用场景将主要集中在借助视频和图像 AI 技术实现自动识别和辅助决策系统。这些领域共同的需求体现在：

- 无线化：包括 MR/AR/摄像头/机器人等多种灵活的无线接入方式；
- 大带宽：高清视频传输需要的网络；
- 确定时延：工业现场毫秒级的确定时延；
- 高算力：基于图像分析的 AI 推理能力、图像渲染需要的计算能力；
- 碎片化：对算力的需求来自各行各业零散需求，存在突发和碎片化的特点；
- 可信：由于要涉及数据的处理，企业对算力处理有可信性，价值可量化、可交易的需求。

目前已经出现的场景，覆盖了运营商 ToB 和 ToC 领域。

2.3.1 运营商 ToB 的“5G 园区+AI”场景

企业建设 5G 园区的主要驱动力来自视频类业务的驱动力，包含监控、智能生产、园区自动驾驶等需要更大带宽和更高实时处理的业务需求。这类业务对边缘智能方案有几个约束：

- 数据隐私要求：原始数据不上公有云；
- 节省带宽成本：不显著增加专线和数据中心的成本，部分原始数据本地分析处理，按需存储在中心 DC；
- 极限边缘部署：室外或 IT 机房场景并存，对噪音、功耗有要求，需要支持无风扇自然散热、更宽的环境温度和低功耗；
- 低时延要求：企业内一些与安全、生产相关的 AI 应用，需要 ms 级的低时延。

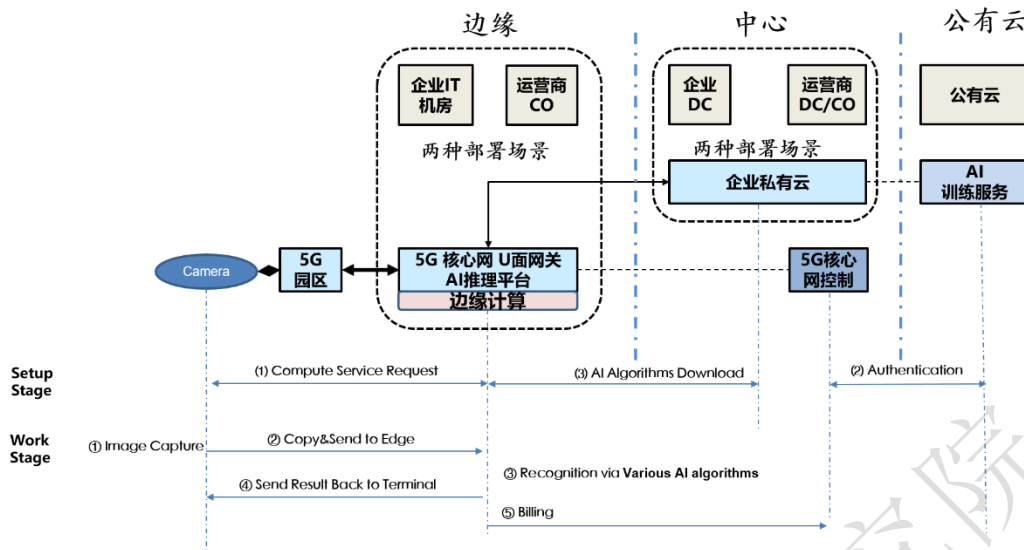


图 2-8 5G 园区+AI 应用场景

在企业边缘，需要在有限的功耗和空间约束下，支持核心网 U 面功能、支持视频 AI 处理，并要考虑将多个边缘的算力共享，组成资源池，最大限度地提升算力的效率。公有云上可以提供一些企业 AI 训练的赋能服务。

2.3.2 运营商 ToC 的“5G+Cloud X”场景

带宽和时延是 Cloud X 业务体验的最大挑战，其中 5G 空口和核心网处理时延相对稳定，传输时延受网络环境（如拥塞），传输距离的影响较大。以 Cloud VR 为例，为保障良好体验，需要端到端时延小于 40ms，要保障极致体验，需要端到端时延小于 20ms，而服务器和终端处理时延已经超过 15ms，要保障极致体验，留给网络“5G+传输+MEC”的总时延只有不到 5ms。

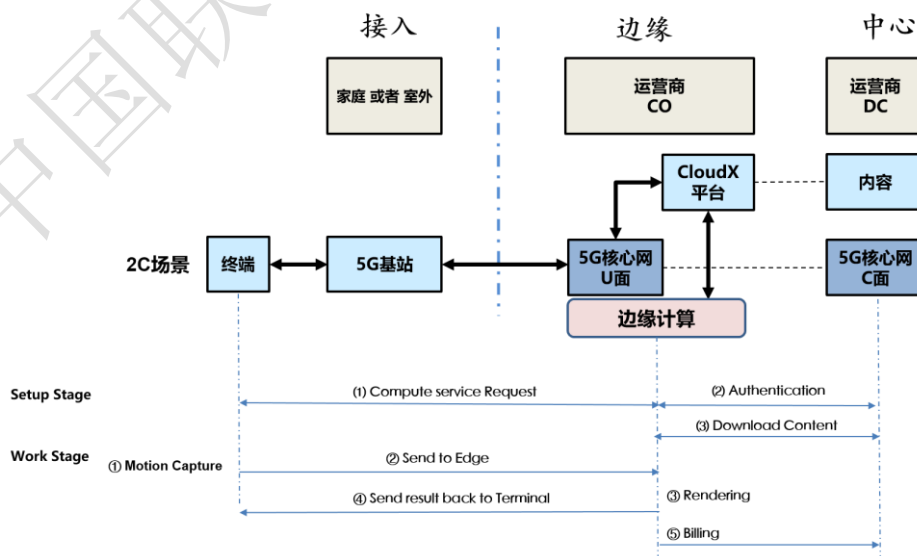


图 2-9 5G + Cloud X 应用场景

为保障业务体验，满足苛刻的时延要求，需要考虑 C/U 分离的核心网架构，U 面下移到运营商边缘，靠近用户提供服务。Cloud X 业务主要需要借助 GPU 的算力，为降低每用户服务成本，需要提高算力的使用效率。为保障业务体验，需要减少网络时延抖动的发生，需要业务和网络之间增加感知和预测能力，在业务发放时，根据网络状态主动做出均衡。

2.3.3 算力开放，运营商提供可交易的算力通证

智能社会里，数以万计的互联网公司是流量的获取者、算力的需求者，而运营商拥有大量的边缘机房和网络云化后边缘算力的优势，是算力的拥有者和流量的引导者。为了适应智能社会的发展，一种动态、快速匹配上述供需关系的新型商业模式是：互联网公司/第三方从运营商按需调用边缘算力，流量获取来自运营商的算力网络，从而节省大量边缘计算重资产投资和流量获取成本；运营商释放边缘机房、网络云化优势，变现“网络连接及其相关的数字资产(计算和存储)”。智能社会带来的这种基于共享经济的新商业模式，可以结合图 2-9 所示的技术方案加以描述：

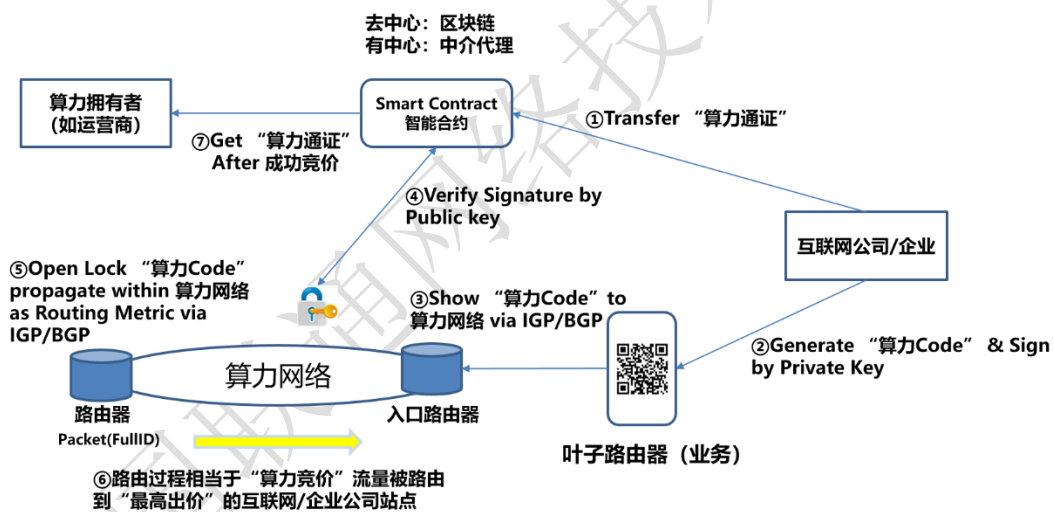


图 2-10 通过算力通证进行算力交易应用场景

需要说明的是，“算力通证”与运营商计算站点可提供的计算能力是等价，即“算力通证”是计算站点可提供的计算能力的一种商业抽象，而“算力 Code”是计算站点可提供的计算能力的一种技术抽象，“算力通证”以“算力 Code”的技术形式在算力网络中传递。

3 算力网络的标准与生态

算力网络涉及多个国际标准组织，如 3GPP、IETF、ETSI、ITU、BBF 等，覆盖了无线网、光传输、数据通信和云核心网相关的技术和架构。3GPP 定义了 5G 大带宽、低时延，核心网 C/U 分离架构的标准，为算力网络奠定了基础，IETF 正在讨论智简网络和计算优先网络的相关标准和协议、ETSI 定义了电信边缘计算 MEC 相关的架构和接口标准，ITU 在全光网、电信可信区块链的领域定义相关标准。由联通、电信和华为共同推动的算力网络顶层架构标准“Framework and architecture of Computing Power Network”在 2019 年 10 月的 ITU SG13 全会上成功立项。由联通和华为共同推动 Metro Computing Network (城域算网)，在 2019 年 8 月份 BBF 标准会议上也已经立项通过。

在一些产业联盟组织里面，如 GSMA、SNAI、IEC、网络 5.0 对于算力网络相关的如确定时延网络、AI for Network、边缘计算等技术的应用场景都有一些讨论。但目前还缺乏一个全面、系统性的对算力网络进行架构和需求定义的组织。

从国家战略与 5G 和 AI 发展的趋势判断，中日韩作为制造业大国，都面临人口老龄化的挑战，必将引领信息化向智能化的产业升级，也将是全球“算力网络”最早研究和实践的地区，邻国日本 2018 年开始，经产省的 IOT 政策已经从工业信息化的“Industry4.1 J”向“Connected Industries”与“Society 5.0”演进，目标行业包含了自动驾驶与出行、制造与机器人、生物技术与材料、成套设备与基础设施维保、智能生活五大行业，希望以“边缘领域”为中心加以实现或强化，将高品质生产经验集中到“无需复杂软件”的边缘计算领域，并进行共享和推广利用，这个思路也符合中国从制造业大国向制造业强国迈进的可持续发展诉求。

云化网络时代，核心需求来自于云计算和 IT 的产业，这个领域美国处于领先地位，因此在云化网络的架构、标准、生态上发挥了主导作用，引领了过去 10 年的技术创新和网络演进。

算力网络时代，需求来自于 5G 与 AI 结合的各行各业智能化的诉求，中国已经在这些领域处于产业第一阵营，结合制造、消费、交通各行业发展的诉求，未来有可能在算力网络的架构、标准和生态上发挥主导作用，为构建全球智能社会做出更大的贡献。

4 总结与展望

运营商掌握两个最重要的物理资源，大量机房局点和覆盖很好的光纤光缆网络，光纤光缆资源伴随着全光接入和带宽提速，潜力被充分发掘，每根光纤从 1G 接入到 10G、100G 接入能力持续提升，可以给运营商带来源源不断的收益，而与此同时，光进铜退和 PSTN 退网，机房空间出现闲置，收益率正在下降。

算力网络为运营商提供了服务升级的可行路径：盘活大量机房局点资源，避免粗放、低层次的出租空间方式，基于算力网络为各行业中对带宽、时延有要求的高价值客户，提供可信、有保障的算网一体的综合服务。

算力的需求来自于智能社会中的各行各业和个人，是一种普遍服务。算力与业务可以解耦，是一种开放服务。算力作为一种新型服务，与网络时延、带宽成本、公众可信力都存在密切联系，这三点运营商都具备明显的资源和能力优势。

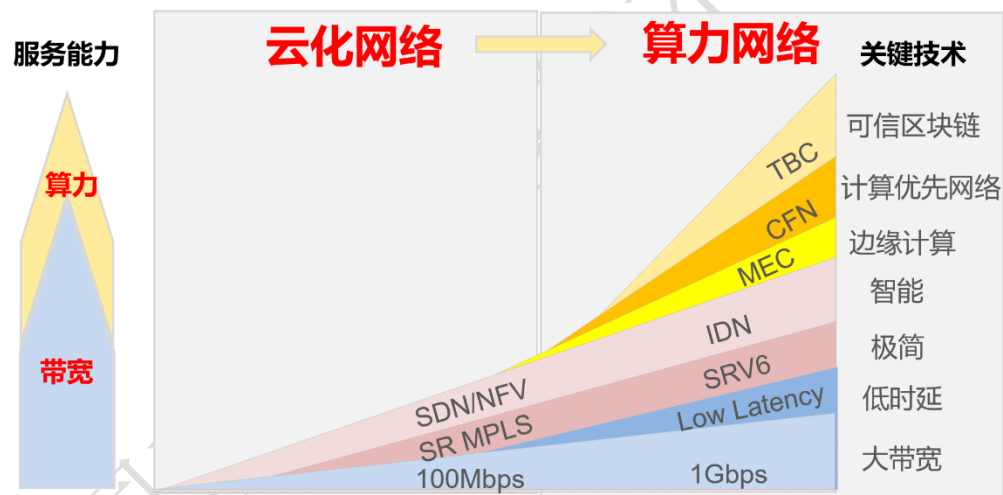


图 4-1 从云化网络到算力网络的技术变迁

对运营商现有网络设备的影响主要体现在城域和接入层设备上：

- 边缘和接入路由器设备需要支持内置计算能力，短期内用于支持网络自动驾驶对运维数据的采集和预处理，长远来看会支持基于 AI 的信号处理算法，实现更大的数据吞吐能力。
- 运营商现有部分 CO 站点需要改造，支持 MEC 的部署，为企业 5G+AI 的应用提供低时延的算力开放平台。
- 运营商现有城域网需要扁平化改造，实现用户接入到业务处理一跳直

达。

- 数据通信协议会增加对数据和算力的感知能力，从云化网络基于 VxLAN/SRv6 服务于租户级联接到算力网络基于 CFN 服务于业务级联接需求。
- 5G/WLAN/以太网物理层会进行优化，支持更低时延和确定性时延的能力，满足垂直行业智能化、自动化的业务诉求。
- 互联网 IP 地址分配和管理机制也将发生变化，从目前全球美国集中控制，走向去中心化、区域化分布式管理，区域之间通过可信电信区块链进行信息交互和授权。

通过新技术的引入和设备的改造，算力网络一定可以作为运营商未来可以经营，并且能够经营好的普遍服务存在，这也是未来网络与计算融合发展的重要方向。

5 缩略语

缩略语	英文全称	中文释义
AI	Artificial Intelligence	人工智能
AR	Augmented Reality	增强现实
BBF	Broadband Forum	宽带论坛
C/U	Control/User	控制/用户
CFN	Compute First Network	计算优先网络
CO	Central Office	中心局
CPU	Central Processing Unit	中央处理器
CT	Computed Tomography	电子计算机断层扫描
CUBE	Cloud-oriented Ubiquitous-Broadband Elastic Network	面向云服务的泛在宽带弹性网
DC	Data Center	数据中心
DCI	Data Center Interconnection	数据中心互联
DCN	Data Communication Network	数据通信网络
DNS	Domain Name System	域名系统
ECC	Edge Computing Consortium	边缘计算产业联盟
ETSI	European Telecommunications Standards Institute	欧洲电信标准化协会
GPU	Graphics Processing Unit	图形处理器
GSMA	Global System for Mobile Communications Association	全球移动通信系统协会
IB	InfiniBand	无限带宽
ICT	Information and Communication Technology	信息和通信技术
IDN	Internationalized Domain Names	国际化域名
IEC	International Electrotechnical Commission	国际电工委员会
IETF	The Internet Engineering Task Force	国际互联网工程任务组
IoT	Internet of Things	物联网
IT	Internet Technology	互联网技术
ITU	International Telecommunication Union	国际电信联盟
M2M	Machine-To-Machine	机器对机器
MBB	Mobile Broadband	移动宽带业务
MEC	Multi-access Edge Computing	多接入边缘计算
MR	Mix Reality	混合现实
NFV	Network Function Virtualization	网络功能虚拟化
OTT	Over The Top	互联网公司越过运营商
P2P	Peer to Peer	对等网络
PC	Personal Computer	个人计算机
PSTN	Public Switched Telephone Network	公共交换电话网络
RAN	Radio Access Network	无线接入网
RDMA	Remote Direct Memory Access	远程直接数据存取
SDN	Software Defined Network	软件定义网络
SNAI	Sensor Network Artificial Intelligence	传感器、网络和人工智能
SRv6	Segment Routing IPv6	IPv6 分段路由

TBC	Trusted Block Chain	可信区块链
TCP	Transmission Control Protocol	传输控制协议
TLS	Transport Layer Security	传输层安全性协议
ToB	To Business	面向企业
ToC	To Customer	面向用户
URLLC	Ultra-reliable and Low-Latency Communication	超高可靠与低时延通信
VM	Virtual Machine	虚拟机
VR	Virtual Reality	虚拟现实
VXLAN	Virtual Extensible Local Area Network	虚拟扩展局域网
WLAN	Wireless Local Area Network	无线局域网