



XXXX

# 面向AI Agent互联网的计算型智能宽带网络网关架构及应用探讨

翁思俊, 颜宇, 杨新国, 宋健, 姜文颖, 程伟强  
(中国移动通信有限公司研究院, 北京, 100053)

**摘要:** 生成式AI与大语言模型推动互联网向AI Agent互联网演进。宽带网络网关(BNG)作为宽带用户接入网络的关键设备,难以适配海量异构AI Agent超大规模管理与边缘算力的新需求。智能宽带网络网关(iBNG)是传统BNG的智能化升级形态,但其仍面临系列挑战。作为前瞻性理论研究,本文面向AI Agent互联网,提出计算型iBNG创新架构,系统阐述其技术架构与关键技术体系,并结合智慧家庭、无人商业、小型工业三大典型场景探讨应用路径。该架构深度融合“联接-计算-智能”能力,通过内生算力与智能管控一体化设计,为AI Agent互联网提供系统性接入网关解决方案。

**关键词:** 计算型iBNG; AI Agent互联网; 宽带网络网关; 计算型网关

**中图分类号:** TP393

**文献标志码:** A

**doi:** 10.11959/j.issn.1000-0801.

## Discussion on the Architecture and Application of Computing Intelligent Broadband Network Gateway for the AI Agent Internet

WENG Sijun, YAN Yu, YANG Xinguo, SONG Jian, JIANG Wenying, CHEN Weiqiang  
China Mobile Research Institute, Beijing 100053, China

**Abstract:** The evolution of the Internet toward the AI Agent Internet was driven by generative AI and large language models. As a key device for broadband user access to the network, the broadband network gateway (BNG) was difficult to adapt to the new requirements of ultra-large-scale management and edge computing power posed by massive heterogeneous AI Agents. The intelligent broadband network gateway (iBNG) was proposed as an intelligent upgrade of the traditional BNG, yet a series of challenges were still faced. As a forward-looking theoretical exploration, this paper proposes an innovative computing iBNG architecture for the AI Agent Internet, systematically elaborates its technical architecture and key technology system, and discusses its application paths in three typical scenarios: smart home, unmanned commerce, and small-scale industry. By deeply integrating the capabilities of ‘connection, computing, and intelligence,’ this architecture provides a systematic access gateway solution for the AI Agent Internet through the integrated design of endogenous computing power and intelligent management and control.



**Key words:** Computing iBNG, Agent Internet, Broadband Network Gateway, Computing Gateway

## 1 引言

生成式AI与大语言模型驱动互联网向智能体(AI Agent)互联网演进, AI Agent成为泛在边缘应用的核心载体<sup>[1-6]</sup>。其规模化部署对网络提出了超大规模异构管理与边缘高强度算力支撑需求<sup>[7]</sup>, 传统连接型网络已难以适配; 随着5G-Advanced向6G演进, 网络亟需向网算智一体转型, 通过融合计算与智能能力, 破解集中式云计算的高时延、数据隐私等瓶颈, 满足AI Agent低时延、高并发算力需求<sup>[7-8]</sup>。

面对上述需求, 边缘智能依托大语言模型与边缘计算的深度融合<sup>[8-9]</sup>, 实现了通信-计算协同优化。其中, SLED框架通过推测解码缓解边缘侧大模型推理的资源瓶颈, 开源模型部署方案则进一步降低部署复杂度<sup>[10-11]</sup>。相关技术进展验证了边缘智能的可行性, 为网络向网算智一体演进、适配AI Agent互联网提供支撑, 并对边缘侧核心接入基础设施提出新的承载要求<sup>[7-11]</sup>。

宽带网络网关(Broadband Network Gateway, BNG)作为用户接入网络的核心锚点与第一跳设备, 具备广泛网络覆盖与天然流量汇聚能力, 是承载边缘算力、实现联接与计算深度融合的理想基础设施<sup>[12-13]</sup>。然而, 传统BNG以连接管理与数据转发为核心能力, 缺乏对异构算力的感知、统一调度与协同能力, 是制约AI Agent部署与高效运行的关键瓶颈。因此, 推动传统BNG向算力与连接融合的智能承载平台加速演进, 已成为支撑算网智一体发展的紧迫课题。已有研究系统指出边缘智能<sup>[8-9,14]</sup>、AI与网络的融合<sup>[15]</sup>以及6G场景下AI Agent网络的趋势<sup>[2-4]</sup>; 产业界也已启动智能型宽带网络网关相关标准研究, 并在技术架构、接口功能等核心维度形成初步共识<sup>[16-20]</sup>; 但针对如何改造传统BNG, 使其原生集成计算、网

络与智能能力以高效、经济地承载AI Agent服务, 尚缺乏系统性的架构设计和技术路径探讨。

作为前沿性理论探索, 本文以上述产业共识为设计基础, 聚焦计算型智能宽带网络网关(intelligent BNG, iBNG), 系统剖析其演进动因与核心挑战, 详细阐述其核心技术架构与关键技术体系, 并结合典型场景探讨其应用路径。上述理论探索为推动边缘网关向智能算力平台演进、支撑AI Agent与边缘智能技术的规模化落地, 奠定了理论与技术基础。

## 2 面向AI Agent的宽带网络业务趋势、需求及挑战

### 2.1 网络业务趋势

随着AI大模型技术突破性进展, 普惠型AI推理成本断崖式下降, AI Agent泛在化部署与具身智能规模化落地进程加速, 全球逐步迈入智能社会新阶段。宽带网络业务呈现三大环环相扣、逐层递进的核心发展趋势<sup>[21-24]</sup>:

趋势一: 基础AI能力普惠化, 推理服务边缘规模化部署。大模型性能迭代提升、应用成本持续降低, 推动AI技术从单一交互工具升级为各行业核心生产力组件, 催生十亿级终端与百万级智能设备的海量高并发实时推理需求, 计算负载从云端向数据源头迁移, 网络边缘成为AI算力部署关键节点。

趋势二: 应用形态向AI Agent化演进, 多体协同成主流。AI Agent从单一应用辅助工具, 进化为具备自主感知、决策与执行能力的独立数字实体, 跨设备、跨场景分布式协同需求激增, 对边缘本地算力的实时数据处理与动态调度能力提出更高要求。

趋势三: AI Agent具身化发展, 物理场景落地加速。人形机器人、工业巡检机器人等具身智

能设备已从实验室研发走向规模化工程应用，高频人机、机机交互对网络时延、可靠性提出严苛标准，且跨厂商设备协议异构性问题日益凸显，对网络生态横向协同能力形成新挑战。

## 2.2 网络能力需求

上述业务趋势驱动宽带网络从传统连接驱动，向“边缘算力+智能内生+生态协同”一体化模式转型，网络流量从被动比特流升级为承载智能交互与价值计算的令牌流。作为固定宽带接入网络的核心枢纽，传统BNG的功能定位与能力体系面临根本性重构：

(1) 管理对象从用户转向海量异构AI Agent。AI Agent化应用与具身智能落地，使传统以人/终端为颗粒度的管理模式已难以适配。传统BNG需建立单个AI Agent精细化管理机制，实现注册、认证、授权、状态监控与注销的全流程自动化管控，并为AI Agent分配唯一Agent ID，按业务类型与服务等级制定差异化策略。

(2) 升级业务支撑能力，形成多AI Agent协同的协议兼容与寻址体系。传统BNG需构建新型业务支撑体系，满足多AI Agent协同和具身智能设备异构性需求，构建新型寻址体系，兼容Matter、MQTT、HL7/FHIR等多种异构通信协议，实现海量异构AI Agent的规模化接入；同时保障AI Agent跨节点迁移后的身份一致性与能力快速匹配，并具备深度业务感知能力，为不同业务提供差异化服务质量保障。

(3) 重构算力供给，从连接管控算力升级为边缘服务算力。普惠型AI推动推理任务从云端下沉至网络边缘，驱使网关升级为“联接+算力”一体化基础设施。通过集成CPU、GPU等异构算力单元，构建规模化异构算力池，将管控算力升级为服务算力，满足AI Agent的实时决策和具身智能的低延迟交互需求。

(4) 构建面向AI Agent的智能化运维与安全体系。面对AI Agent化和具身智能带来的规模复

杂度与安全挑战，传统BNG的运维与安全能力亟需升级。运维侧需建立AI Agent级业务感知能力，关联网络KPI与业务KQI，通过AI运维实现故障预测、诊断与自愈；安全侧应构建以AI Agent数字身份为核心的零信任安全体系，通过持续认证、端到端加密等机制确保AI Agent交互安全可靠。

## 2.2 计算型iBNG的技术演进挑战

AI普惠、AI Agent泛在化与AI Agent具身化的深度叠加，推动宽带网络向“边缘算力+智能内生+生态协同”转型，也对传统BNG向计算型iBNG演进提出新挑战：

(1) 边缘算力供需与成本效率失衡。边缘侧需承接十亿级用户与百万级AI Agent的并发计算任务，算力需求呈指数级增长；同时，不同场景时延要求从毫秒级到亚毫秒级差异显著，算力规模与时延性能面临双重压力<sup>[25-26]</sup>。此外，由于边缘侧缺乏高效本地化数据处理方案，易造成带宽资源浪费与云端成本高企，在有限资源下的算力供需与成本平衡成为首要挑战。

(2) 生态协同缺失与协议适配困难。AI Agent生态呈现多厂商、多协议、多标准并存的异构特征，跨厂商、跨领域AI Agent间缺乏直接通信能力。传统BNG作为网络侧的核心协同枢纽，目前缺乏统一的协议适配标准与协同框架，难以实现异构AI Agent间的互联互通，制约其应用规模化发展，这也是当前产业界制定宽带网络网关标准的核心攻关方向<sup>[16-20]</sup>。

(3) 设备技术重构与演进难度大。传统BNG围绕连接管控设计，无法满足AI Agent互联网需求。向计算型iBNG演进需实现异构算力池集成、100G+超宽接入、纳秒至毫秒级确定性时延等技术重构，软硬件升级复杂度较高，对设备研发与网络部署构成现实挑战。

(4) 安全治理滞后与管理模式转型困难。数据本地化闭环是隐私保护与合规要求的刚性需



求，当前边缘侧尚未形成统一的数据隐私保护和模型内生安全防护机制。AI Agent 具备快速创建、频繁迁移、动态销毁等特征，使管理维度大幅增加，运营商现有基于用户/设备的运营体系难以有效适配，管理模式转型面临挑战。

### 3 计算型 iBNG 技术架构与组网

本章从理论设计层面提出计算型 iBNG 的“三层四模块”架构，通过明确边缘算力部署形态、界定各层次逻辑边界、梳理核心技术功能与实现路径，系统构建其整体技术架构，并在此基础上探讨四级泛在智能组网方案。

#### 3.1 计算型 iBNG 设备形态

计算型 iBNG 按边缘算力部署形态分为嵌入式、集群与池化三种类型，以适配不同场景的性能、规模与灵活性要求，三种计算型 iBNG 形态比较如表 1 所示。

**嵌入式形态：**通过在传统 BNG 设备槽位中插入高性能算力板卡，实现网络与计算功能的硬件级紧耦合。该架构能为 vCPE、AI Agent 等提供极致的低时延与电信级高可靠性，适配当前阶段工业控制、在线游戏等需要微秒级响应的深边缘及企业园区场景。在初期，深边缘算力需求较小，可采用嵌入式架构快速构建边缘算力，推动边缘业务的快速落地部署；但其算力规模受物理槽位和单板性能的严格限制，扩展性较差，难以支撑大模型等重算力应用。

**集群形态：**通过高速互联技术将多个设备框组成逻辑集群，实现算力弹性横向扩展，适配网

络功能、云应用、AI Agent 服务等业务的分布式部署，适合流量与数据本地闭环的企业场景，可部署 Agent GW、云化 iCPE 等应用。该架构即解决嵌入式形态算力不足问题，又仅需协调 CT 网络稳态性与 IT 应用敏捷性，效避了池化形态对 IT/CT 跨领域深度融合的强需求。

**池化形态：**算网超融合深度演进方向，通过高速交换互联将异构计算、存储等资源聚合为统一可动态调度的资源池，弹性扩展能力强，适配未来从基础网络服务到大模型训练推理的多样化业务，但其实现需满足 IT 与 CT 深度融合及电信级高可靠性要求。

#### 3.2 计算型 iBNG 的架构设计

为实现“联接-算力-智能”三位一体的核心定位，适配异构 AI Agent 泛在接入与协同需求，本文创新性提出计算型 iBNG “三层四模块”的逻辑架构，打破传统网关重联接、轻算力、弱智能的局限，兼顾通信设备的高可靠性与 AI 业务敏捷迭代需求，为 AI Agent 全生命周期管理与跨场景协同提供一体化技术支撑。

图 1 为计算型 iBNG “三层四模块”整体逻辑架构拓扑图，整体架构以分层定界、模块协同为设计原则，纵向分为基础设施层、Agent 网关层、应用层，横向涵盖异构资源模块、网络与智能体管控模块、安全与运维模块、弹性扩展模块四大核心模块。其中，基础设施层为底层支撑，Agent 网关层为核心管控，应用层为场景赋能，四大模块嵌入各层级实现网络传输、安全防护等功能落地；层级间通过标准化接口交互，模块间

表 1 三种计算型 iBNG 形态比较

部署形态	部署位置	算力供给特征	核心技术功能	适用场景
嵌入式	区县、企业园区边缘节点	<b>小规模算力。</b> 算力本地集成，规模适配单场景轻量需求	轻量算力供给、微秒级低时延转发、单场景 Agent 接入管控	Agent GW、云化 iCPE 等单场景边缘业务。
集群	区县、企业园区边缘节点	<b>中规模算力。</b> 算力横向扩展，支撑多场景算力协同需求	弹性算力调度、多协议兼容适配、跨场景 Agent 协同	Agent GW、云化 iCPE、中轻量边缘云应用
池化	地市/大区核心边缘节点	<b>大规模算力。</b> 算力资源池化聚合，弹性适配多样化重算力需求	全域算力调度、大模型推理支撑、跨域 Agent 全局协同	Agent GW、云化 iCPE、大模型推理等重算力边缘云应用



图1 计算型iBNG“三层四模块”整体逻辑架构图

协同联动保障端到端服务能力；其三层架构组件与四模块的核心技术特性将分别在3.2.1与3.2.2节展开阐述。

### 3.2.1 三层架构组件

三层架构作为图1所示架构的纵向结构，依托横向四模块实现全维度功能支撑，共同完成多/异构AI Agent接入与跨场景协同。其中，

基础设施层位于架构最底层，通过对计算、存储等物理资源进行虚拟化与池化，构建统一的虚拟资源池；基于标准化接口实现资源的按需编排与动态调度，从而为上层应用提供稳定、高效的基础资源服务。

Agent网关层作为核心管控层，是实现向智能服务跃迁的关键层级。该层集成AI Agent管理组件，完成异构AI Agent的接入认证、状态监控与全生命周期管控；通过精细化的业务感知，识别AI Agent类型与SLA需求，适配高频短对话、长连接维持等流量特征，并联动DNS等基础服务实现标准化网络寻址与资源分配。同时，融合家庭及边缘算力资源，为AI Agent提供本地化推理支撑与长期记忆能力，辅以网络质量自感知与动态调优，从而为端到端的AI Agent服务提供稳定保障。

应用层聚焦AI Agent的场景化部署与协同应用，基于底层资源与管控能力，面向养老、医疗、健康等领域部署场景化AI Agent。各AI Agent通过标准化接口与网关层交互，实现资源按需调用与跨AI Agent协同，形成“资源供给-智能管控-场景应用”的闭环体系，充分释放计算型iBNG“联接-算力-智能”三位一体的架构优势。需明确的是，应用层本身不包含四大模块的具体组件，但其运行完全依赖于四大模块在算力、网络、安全与运维、弹性四方面提供的全域支撑。

### 3.2.2 四大核心模块技术特性

四大核心模块是图1中架构的横向能力载体，深度嵌入基础设施层、Agent网关层各环节，为应用层提供全流程能力支撑，各模块既具备专业化的核心技术特性，又通过协同调度机制实现能力联动，共同支撑异构AI Agent的可信接入、高效协同与场景化赋能。

(1) 异构资源模块是计算型iBNG的核心算力与存储单元，为AI Agent业务提供异构算力供给、虚拟化调度及多类型任务处理能力。模块支持CPU、GPU、NPU混合部署，可根据业务特征提供差异化算力服务；配套存储单元承担数据持



久化、容量扩展与高可靠存储，为AI Agent上下文记忆与模型缓存提供支撑。在嵌入式及集群形态下，模块采用“板卡式集成+模块化部署”设计，算力卡通过高速接口与主板互联，保障低时延高带宽算力交互，同时集成算力虚拟化引擎实现资源精细调度；在池化部署形态下，支持类数据中心无损部署，实现算力资源全局池化与动态调度。

(2) 网络与智能体管控模块是计算型iBNG的核心能力单元，纵向贯通基础设施层与Agent网关层，为应用层提供网络传输与Agent管控支撑，具备超宽接入、差异化QoS、异构AI Agent精细化管理等能力。在网络传输方面，模块针对多AI Agent协同流量特征，集成高性能可编程转发芯片，优化转发路径与缓存机制，实现亿级并发连接快速转发，保障南北向推理流量与东西向协同流量的高可靠传输；支持SRv6/G-SRv6等协议，实现与城域边缘网及骨干网的无缝对接；具备层次化网络切片与QoS优先级调度能力，为工业推理、应急响应等核心业务提供确定性带宽与时延保障。在智能体管控方面，模块为海量异构AI Agent提供动态可信接入与精准交互能力，通过身份认证、语义解析、能力寻址实现从位置可达向能力匹配的服务升级；结合动态授权控制与权限全生命周期管理，实现AI Agent精细化管理；依托全维度信息采集与智能分析，搭建业务感知调度与体验保障体系，为AI Agent跨域协同与场景化应用提供核心支撑。

(3) 安全与运维模块是计算型iBNG的内生安全保障与运维管理单元，具备零信任接入、攻击防护、智能运维等功能，为架构稳定、安全、高效运行提供基础支撑。在内生安全保障方面，模块围绕AI Agent接入与算力服务构建全流程防护能力：集成国密二级安全芯片，支持数据加密、身份认证与密钥管理；内置DDoS攻击防护单元，基于硬件加速流量清洗技术，快速识别并

拦截TCP SYN Flood、UDP Flood等常见攻击；新增AI模型加密模块，对部署于算力集群的模型进行硬件级加密存储与授权调用；配备硬件级可信根，保障设备启动过程安全。在运维管理方面，模块配备专用管理接口，支持远程运维与设备状态实时监测；预留本地存储扩展槽，用于日志存储、模型缓存与应急备份；依托AI技术实现预测性维护、故障快速定位与差异化处置、算力协同优化，提升运维自动化水平与资源利用率。

(4) 弹性扩展及运维模块是计算型iBNG的资源弹性扩展与应急保障单元，核心具备集群/池化扩展、资源弹性调度与本地应急保障等功能，支撑架构向集群化、池化形态演进。集群形态下，模块配置高速集群互联接口，实现多设备堆叠组网与算力资源横向扩展；池化形态下，配置服务器与中心网络设备高速互联接口，实现多服务器池化组网与算网资源全局弹性调度。模块可与异构资源模块、网络与智能体管控模块协同，根据AI Agent业务流量与算力需求动态调整资源供给，保障业务连续性与稳定性；配套本地应急保障能力，可在网络中断、算力故障等异常场景下实现核心业务本地降级运行，提升架构容错能力与可靠性。

### 3.3 面向异构AI Agent协同的泛在智能组网

基于异构AI Agent的泛在接入与跨域协同需求，计算型iBNG突破传统BNG仅承载网络联接的功能局限，以“三层四模块”技术架构为核心载体，构建端-边-网-云四级泛在智能组网体系，实现算网资源协同优化与跨域高效协同。图2为面向AI Agent协同的四级泛在智能组网示意图，该层级涵盖终端、边缘、网络协同、云端支撑四个组网层级的核心内容，各组网层级层层递进、能力叠加，可根据业务场景与协同需求灵活适配。

终端算力级是泛在智能组网的基础层级与异构AI Agent接入口，组网形态最为小型化，主

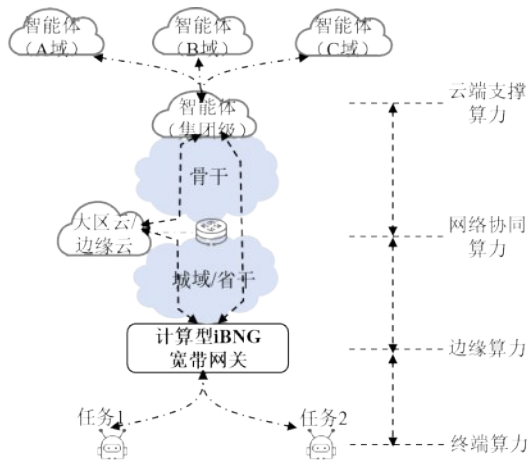


图2 面向 AI Agent 协同的四级泛在智能组网示意图

要涉及智能家居、安防传感等终端 AI Agent。该层级通过计算型 iBNG 标准化接口实现异构 AI Agent 统一接入，向边缘算力级上报自身能力特征，接收并执行边缘下发的本地轻量型任务。

边缘算力级是应用最广泛的核心层级，在终端算力级基础上新增计算型 iBNG 设备，集终端管控、业务适配、云边衔接于一体。计算型 iBNG 作为核心载体，一方面通过身份认证与能力注册实现家庭侧 AI Agent 统一管控，完成异构 AI Agent 可信接入与能力登记；另一方面提供终端数据就近处理、内生算力调度、低时延响应等能力，依托内生算力实现终端数据本地化解析，为时延敏感业务提供毫秒级端到端响应。同时，通过算网协同能力，计算型 iBNG 有效衔接边缘与云端资源，为泛在智能业务规模化落地提供核心技术底座。

网络协同算力级是泛在智能组网的跨域协同层级，在终端与边缘能力基础上新增跨域 AI Agent 协同能力。该层级基于计算型 iBNG 资源调度能力，结合网络资源动态适配与差异化 QoS 调度，实现跨设备、跨场景、跨地域 AI Agent 高效协同，为跨域场景应用提供算网协同支撑。

云端支撑算力级是泛在智能组网的全维度层级，集成前三个层级所有能力，同时新增算力能

力扩展、全局调度联动、业务能力开放三大能力，是面向海量异构 AI Agent 协同的终极组网形态。该层级中，计算型 iBNG 对接省/集团级边缘云平台，具备重算力业务扩展接口，可将重算力型业务卸载至云端处理，解决边缘算力不足问题；同时支撑跨区域 AI Agent 全局调度与跨域业务联动，实现端-边-网-云算力资源协同优化。

### 4 计算型 iBNG 面向 AI Agent 的关键技术

计算型 iBNG 作为承载边缘算力与网络接入的核心设备，是衔接异构 AI Agent 与广域智能体网络的关键设施，支撑网络使能智能体与智能体赋能网络的双向赋能。基于“三层四模块”架构与四级泛在智能组网，本章从理论设计维度系统阐述计算型 iBNG 的智能连接与可信管理、感知调度与体验保障、内生安全与自主运维三大关键技术体系，旨在为设备落地及边缘算网融合规模化部署提供技术支撑，并为后续应用场景探讨奠定理论基础。

#### 4.1 智能连接与可信管理技术

智能连接与可信管理技术是实现海量异构 AI Agent 动态接入与可信交互的基础，核心解决身份统一标识、动态组网互通、任务级授权管控三大问题，突破传统以用户/IP 为核心的连接管理模式，构建统一标识、动态寻址与高效管控的技术体系，支撑“人-设备-AI Agent”一体化协同交互。为此，本文从以下三个维度构建智能连接与可信管理技术体系。

在统一身份识别方面，采用“网络-身份”两层复合标识，实现对异构 AI Agent 的全局唯一识别与跨域互信。网络层沿用 IPv6 全局寻址，保障 AI Agent 基础网络可达，满足物理终端传统接入需求；身份层引入数字身份标识 Agent ID，并通过区块链分布式账本记录身份信息与能力属性，实现跨平台互信与行为可追。同时，Agent ID 同步关联用户身份、设备标识及算力水平、功



能特性等能力属性，构建“身份-能力-权限”关联机制，为后续寻址、管控、安全防护提供基础。

在动态寻址方面，构建“IP地址寻址-身份寻址-能力寻址”三级协同机制，实现从位置可达到能力匹配的精准服务发现，其具体如图3所示。基于统一身份标识，计算型iBNG作为边缘侧能力注册节点，同步AI Agent身份信息、能力特征与网络位置；同时，通过AI Agent DNS新增能力资源记录，支持基于业务意图的能力检索与地址映射。其核心流程为：IP地址验证网络可达性，Agent ID完成身份真实性认证，能力匹配度评估，服务实例可用性确认。该机制既支持已知AI Agent ID的高效路由，也能在未知目标场景下实现从业务意图到能力匹配的服务发现。

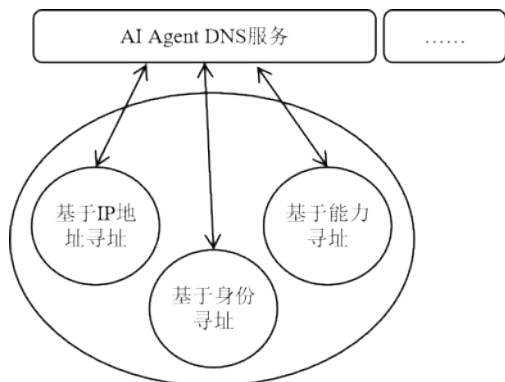


图3 AI Agent动态寻址机制示意图

在高效管控方面，从会话管理、动态授权、全生命周期管理三方面构建精细化管控体系，以解决传统网络管控颗粒度粗、授权静态、高并发承载弱的问题。会话管理采用状态分级优化策略，活跃会话存于高速内存池保障处理效率，非活跃会话经压缩后持久化存储，恢复时可亚秒级唤醒，提升系统高并发承载能力，适配AI Agent海量并发接入特征。动态授权基于属性访问控制模型，综合AI Agent身份、属性、环境及算力资源动态生成访问策略，权限随任务启动自动激活、结束实时回收，实现一次协商动态适配；同

时，策略随AI Agent移动迁移，减少冗余授权下发，实现一次认证全域通行。全生命周期管理实现AI Agent从注册、发现、监控到注销的全流程自动化，支持按终端类型、业务等级实施分级管控，为具身智能等高优先级AI Agent提供专属接入通道，将管理颗粒度从用户/终端细化至单个AI Agent，保障差异化服务需求。

#### 4.2 感知调度与体验保障技术

在AI Agent互联网中，不同类型Agent对网络和算力需求差异显著，如具身智能Agent要求亚毫秒级时延，大模型推理Agent要求高算力持续供给，传感Agent要求低功耗小流量传输；同时，AI Agent流量兼具显性交与隐性工具调用双重负载特征。传统调度技术以静态带宽分配、通用QoS调度为核心，仅采集链路利用率、丢包率等基础网络指标，缺乏对算力状态与业务体验的协同感知，仅能适应音视频等标准化流量，难以支撑上述新型流量需求。

感知调度与体验保障技术是实现AI Agent业务端到端保障的核心，针对AI Agent流量异构化、需求差异化、算网协同化等特征，突破传统网络仅聚焦带宽分配的单一调度模式，构建“全维度信息采集-智能融合分析-动态资源适配”技术体系，实现网络与算力资源的协同感知与动态调度，保障不同类型AI Agent业务的差异化QoS需求，其技术实现如图4所示。

全维度信息采集构建“网络-算力-业务-环境”四位一体监测体系，实现AI Agent业务运行全链路状态精准捕捉。网络侧采集时延、抖动、丢包率、带宽利用率等传输参数，精准反映连接状态；算力侧采集CPU/GPU/NPU负载、算力利用率、任务排队时长等指标，实时掌握边缘算力供给；业务侧采集指令响应时间、推理成功率等AI Agent专属指标；环境侧采集位置、移动速度、网络负载分布等场景参数，为动态调度提供上下文支撑。

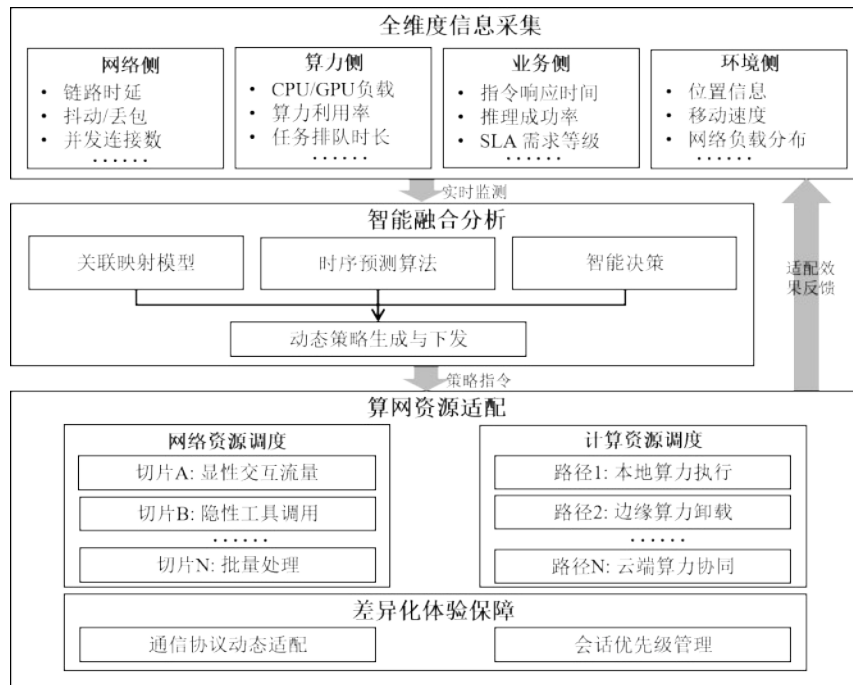


图4 感知调度与体验保障技术实现示意图

智能融合分析作为决策核心，基于网络大模型实现多维度指标关联映射与智能决策。通过机器学习实施“网络-算力-业务”智能决策，量化评估不同类型 AI Agent 的服务质量。例如，针对具身智能低时延需求，建立时延-算力-带宽关联模型；针对大模型推理高算力需求，构建算力负载-任务优先级调度模型；同时，结合时间序列预测，对网络拥塞、算力瓶颈等问题提前预判，为资源适配提供前瞻性决策依据，突破传统人工阈值判断局限。

算网资源适配依托统一资源编排器，实现网络与算力资源的协同动态调度。网络资源调度借助 SRv6 可编程能力，为不同类型 AI Agent 分配专属网络切片，将不同 SLA 要求的业务流导入对应切片，并对显性交互与隐性工具调用流量进行统一标识与同等级保障；计算资源调度通过统一资源编排器，基于任务类型与实时算力状态，动态选择本地执行、边缘卸载或云端协同等模式，实现算网资源联合最优分配；在差异化体验保障中，针对单 AI Agent 多会话资源竞争，明确会话

优先级并进行精细化资源规划，支持 AI Agent 根据流量特征协商适配通信协议，实现资源高效利用。

### 4.3 内生安全与自主运维技术

在 AI Agent 互联网中，AI Agent 的动态性、自主性引入了身份伪造、恶意提示词操控等新型安全风险，海量接入与复杂算网协同对网关故障预判与快速处置能力提出更高要求。传统 BNG 以网络边界防护为核心，采用静态防火墙、基础加密等技术，防护粒度粗且缺乏对 AI Agent 专属数据的安全防护；且其传统运维以人工巡检、事后处置为主，效率低、响应慢，难以适配 AI Agent 海量部署带来的运维复杂度。为此，本文针对 AI Agent 身份动态、交互路径可变、数据密集且流量不可预测等特征，以及算力升级后网关运维复杂度大幅提升的问题，构建“内生安全防护-自主智能运维”一体化技术体系，突破传统边界防护与人工运维局限，支撑计算型 iBNG 的安全高效运行。

图 5 展示了内生安全技术架构，其以智能安

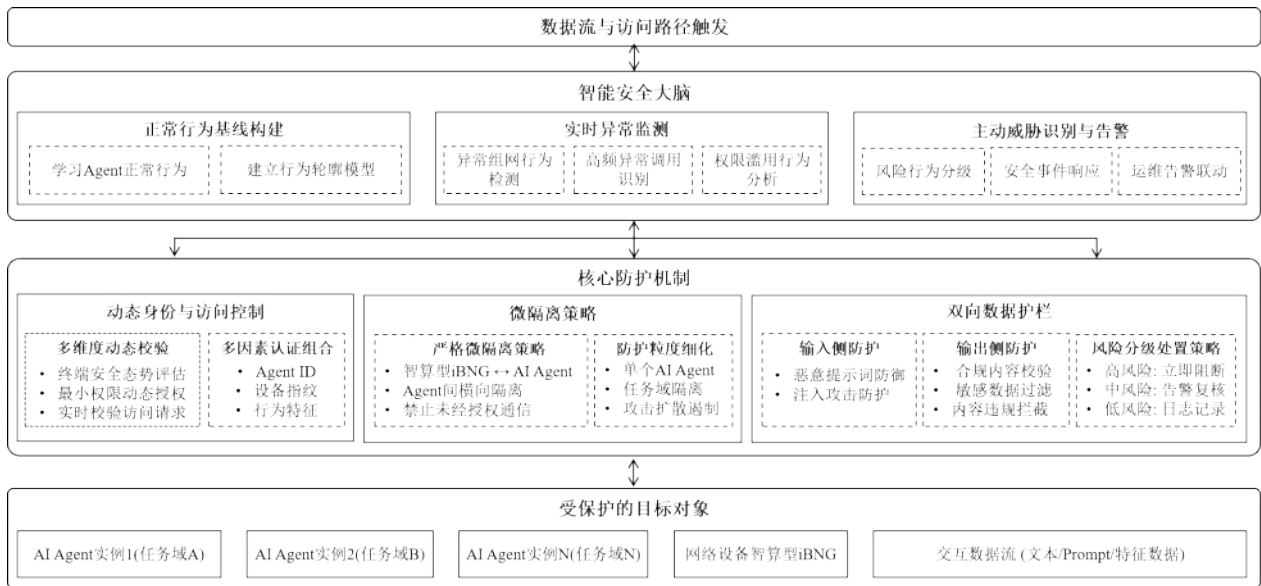


图5 计算型iBNG内生安全技术架构示意图

全大脑和零信任防护机制为核心，由数据流与访问路径触发，对受保护目标对象实施防护。内生安全防护技术以零信任机制为核心，遵循永不信任、持续验证原则，构建全流程安全防护体系，实现从边界防护向全维度防护转变。在动态认证方面，建立多因素认证机制，综合 Agent ID、设备指纹、行为特征对访问请求进行多维度动态校验。在微隔离方面，于计算型 iBNG 及 AI Agent 间实施严格隔离策略，禁止未经授权横向通信，将防护粒度细化至“单个 AI Agent + 任务域”，有效遏制攻击扩散。在数据安全方面，构建双向数据护栏系统，通过输入侧恶意提示词防御、输出侧合规内容校验及分级处置策略，对 AI Agent 交互的文本、prompt 等专属数据实施双向实时监控，精准抵御恶意操控、越权访问、数据泄露及内容违规等风险。此外，在行为审计方面，基于 AI 算法建立 AI Agent 正常行为基线，实时监测异常组网、高频调用、权限滥用等风险行为，实现安全威胁主动识别与告警，将内生安全纳入日常运维。

图6为自主运维技术架构示意图，其由数据底座与采集、自主运维能力、智能运维大脑和可

视化监控四部分构成，共同协作完成对受管目标的运维管理。自主运维技术依托 AI 实现从事后被动处置向主动预判与自动化处置的转变，适配算力升级后复杂的网关运维需求，主要包括预测性维护、故障快速定位与差异化处置、算力协同优化三大功能。预测性维护融合无监督学习与监督学习，对设备日志、性能指标、流量模式等数据进行深度分析，实现硬件故障风险与算力资源瓶颈的预测。故障发生时，借助多段定位、因果推断与知识图谱，快速定位故障根本原因；针对已知故障自动执行预定义修复流程，实现秒级自愈；复杂故障自动生成诊断报告与修复建议，提升运维效率。算力协同优化基于业务负载与算力需求规律，对异构算力资源进行动态调度与负载均衡，提升资源利用率，同时实现设备运行状态可视化监控，降低人工运维成本。智能运维大脑基于采集数据，利用监督学习、融合分析与故障特征库等技术，为自主运维构建算法模型与知识底座，并通过可视化监控实时呈现网络业务状态。

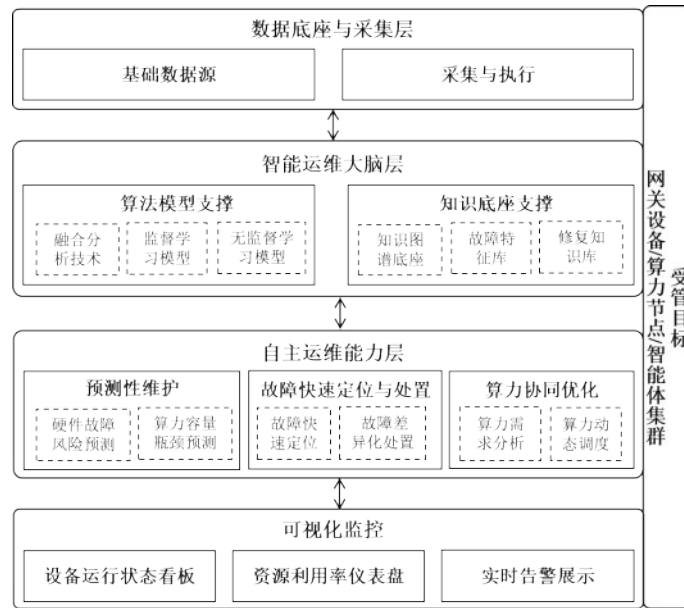


图6 计算型iBNG自主运维技术架构示意图

### 5 计算型iBNG的典型应用

#### 5.1 智慧家庭

家庭数字化正从设备联网基础时代迈向“AI Agent协同+普惠AI推理”的高阶智能时代，核心需求是实现“人+AI Agent”多元服务协同，通过普惠AI推理与软硬件多AI Agent融合，支撑主动预判、无感触发、个性化定制的家庭服务重构。

基于计算型iBNG的智慧家庭整体架构如图7

所示。其中，通过部署家庭管家、康养服务、安防监控等多类型AI Agent，依托计算型iBNG的边缘算力、多阶段寻址和感知调度等能力，构建涵盖日常起居、健康监测和应急响应的全流程服务体系。日常场景中，各类AI Agent通过计算型iBNG实现基于用户作息习惯的联动适配，支撑全天候智能推理与调节；康养场景中，穿戴设备实时采集生命体征数据并开展本地化分析，实现低时延服务与数据不出户；应急场景中，安防摄像头与毫米波检测设备实时采集数据，经计算型

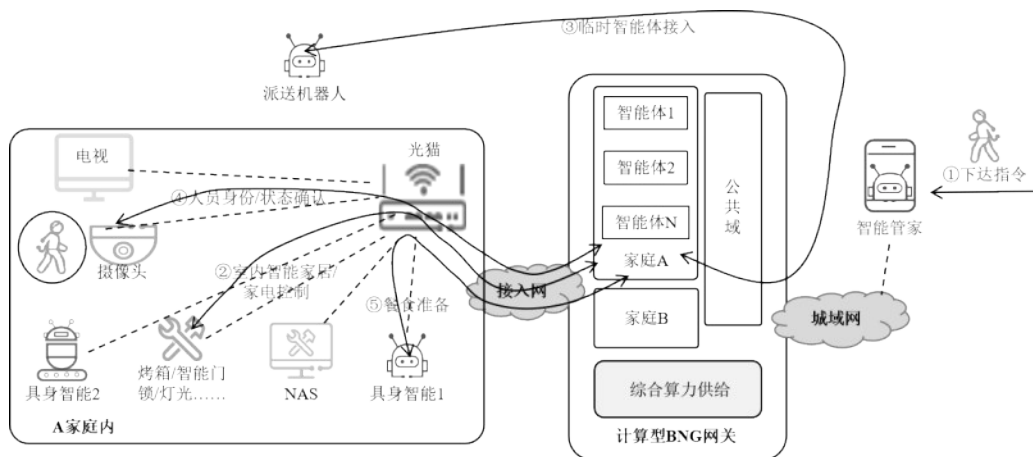


图7 基于计算型iBNG的智慧家庭整体架构示意图



风险。安防 AI Agent 同样依托边缘算力与算网协同，实现安全隐患预判与违规操作识别，筑牢厂区安全防线。物料调度 AI Agent 与生产 AI Agent 协同实现物料动态感知及补给流程自动触发，其运行流程如图9所示。通过实时物料监控与状态分析，系统可对物料短缺情况进行智能预警，并自动生成物料补给请求；中枢 AI Agent 承接该请求后，完成全局物料调度决策与执行操作，最终实现物料精准补给及库存状态的动态更新。此外，计算型 iBNG 依托 SRv6 可编程切片技术，为生产控制信号、视频监控等不同类型流量分配专属传输通道，保障生产控制的确定性与厂区安全的可靠性。

计算型 iBNG 弥补了传统 BNG 在边缘算力、确定性传输、智能协同与安全隔离方面的不足，为小型工业场景提供一体化智能支撑。通过生产流程智能协同与设备预测性维护，提升生产效率与设备利用率，降低运营成本；借助实时安防监测与分级告警，减少安全事故概率，保障人员与财产安全；依托标准化架构与灵活部署，降低中小微企业智能化转型门槛，推动工业互联网向基层场景延伸。

## 6 结束语

AI Agent 互联网背景下，宽带网络面临一系

列核心技术挑战。为此，本文提出融合“联接+算力+智能”的计算型 iBNG 架构。该架构继承传统网关核心定位，通过“三层四模块”实现从用户接入向 AI Agent 业务服务的职能升级，弥补传统 BNG 重联接、轻算力的不足，为 AI Agent 规模化部署提供架构支撑。作为前瞻性理论设计研究，本文围绕该架构阐述了计算型 iBNG 关键技术体系，并结合智慧家庭、无人商业等场景分析其应用路径，从业务逻辑层面验证了技术体系的可行性与应用价值。受限于实验条件，本文暂未开展系统性原型验证，后续将搭建试验环境，围绕三级协同寻址、感知调度、内生安全等关键技术开展实测评估，推动理论架构向工程化应用转化。

结合本文研究内容，计算型 iBNG 的中长期技术演进可重点聚焦三大方向：一是深化 AI 与网络技术融合，提升意图识别精度与资源调度智能化水平；二是推进跨厂商 AI Agent 交互接口与流程规范化，构建开放协同的产业生态；三是探索算力与网络服务一体化计费模式，拓展产业商业价值。

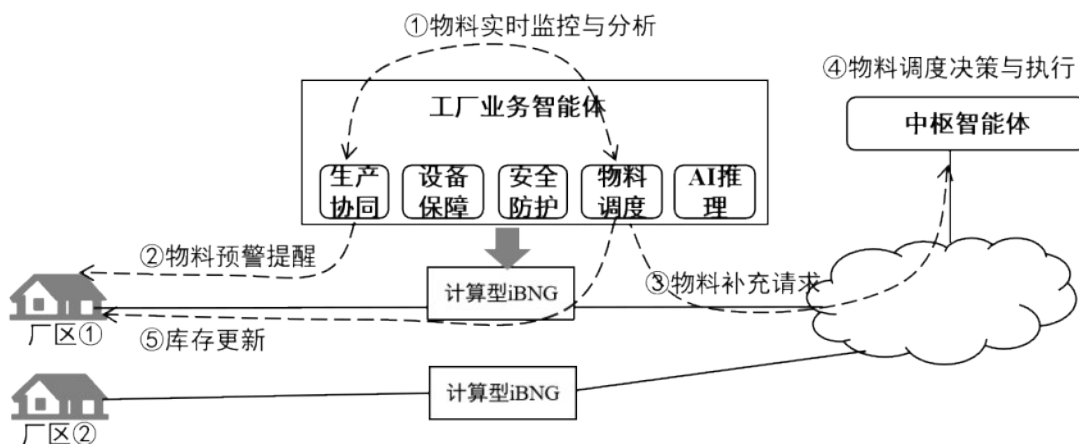


图9 小型工业智能网络架构与物料调度场景示意图



## 参考文献:

- [1] ZHANG R, LIU G, LIU Y, et al. Toward Edge General Intelligence with Agentic AI and Agentification: Concepts, Technologies, and Future Directions[J]. 2025. DOI: 10.1109/COMST.2026.3651702.
- [2] XU M R, NIYATO D, KANG J W, et al. When large language model agents meet 6G networks: perception, grounding, and alignment[J]. IEEE Wireless Communications, 2024, 31(6): 63-71.
- [3] LI B R, LIU T E, WANG W L, et al. Agent-as-a-service: an alternative edge computing framework for 6G networks[J]. IEEE Network, 2025, 39(2): 44-51.
- [4] XIAO Y, SHI G, ZHANG P. Towards agentic AI networking in 6G: a generative foundation model-as-agent approach[EB/OL]. (2025-03-01)[2025-02-11]. arXiv:2503.15764.
- [5] HOU X, ZHAO Y, WANG S, et al. Model context protocol (MCP): landscape, security threats, and future research directions[EB/OL]. (2025-03-01)[2025-02-11]. arXiv:2503.23278.
- [6] TUPE V, THUBE S. Demonstrating multi-agent collaboration via agent-to-agent and model context protocols: an IT incident response case study[C]//Proceedings of 2025 IEEE International Conference on Service-Oriented System Engineering. Piscataway: IEEE Press, 2025: 1-5.
- [7] ZHOU Z, CHEN X, LI E, et al. Edge intelligence: paving the last mile of artificial intelligence with edge computing[J]. Proceedings of the IEEE, 107(8): 1738-1762.
- [8] DENG S, ZHAO H, FANG W, et al. Edge intelligence: the confluence of edge computing and artificial intelligence[J]. IEEE Internet of Things Journal, 2020, 7(8): 7457-7469.
- [9] FRIHA O, FERRAG M A, KANTARCI B, et al. LLM-based edge intelligence: a comprehensive survey on architectures, applications, security and trustworthiness[J]. IEEE Open Journal of the Communications Society, 2024, 5: 5799-5856.
- [10] LIN J, WANG H, CHEN Y, et al. SLED: A Speculative LLM Decoding Framework for Efficient Edge Serving[C]//SEC 2025: Tenth ACM/IEEE Symposium on Edge Computing. Piscataway: IEEE Press, 2025: 1-12.
- [11] 王睿, 张留洋, 高志涌, 姜彤云. 面向边缘智能的大模型研究进展[J]. 计算机研究与发展, 2025, 62(10): 2565-2582.
- [12] KHAN L U, YAQOOB I, TRAN N H, et al. Edge computing enabled smart cities: a comprehensive survey[J]. IEEE Internet of Things Journal, 2020, 7(10): 10200-10232.
- [13] PORAMBAGE P, OKWUIBE J, LIYANAGE M, et al. Survey on multi-access edge computing for internet of things realization[J]. IEEE Communications Surveys & Tutorials, 2018, 20(4): 2961-2991.
- [14] LI M, ZHANG Y, WANG Q, et al. Power internet of things networked group intelligent sensor networking[J]. Journal of Physics: Conference Series, 2021, 1744(1): 012345.
- [15] KATO N, FADLULLAH Z M, TANG F, et al. The deep learning vision for heterogeneous network traffic control: proposal, challenges, and future perspective[J]. IEEE Wireless Communications, 2017, 24(3): 146-153.
- [16] 工业和信息化部. 智能化宽带网络网关(iBNG)管控系统技术要求:2024-0500T-YD [S]. 北京:中国标准出版社,2025.
- [17] 工业和信息化部. 智能化宽带网络网关(iBNG)架构及总体技术要求:2023-11087176-YD [S]. 北京:中国标准出版社, 2024
- [18] BROADBAND FORUM. DBNG for Wired Access (WT-487) [S]. Brussels:Broadband Forum, 2024.
- [19] BROADBAND FORUM. Broadband Forum. Multi-Service Disaggregated BNG with CUPS: Reference Architecture, Deployment Models, Interface, and Protocol Specification (TR-459 Issue 3) [S]. 2025.
- [20] BROADBAND FORUM. Multi-Service Disaggregated BNG with CUPS: Integrated Carrier Grade NAT Function (TR-459.2 Issue 2) [S]. 2025.
- [21] KATENDE R. Rethinking data-efficient artificial intelligence for low-resource settings[J]. Machine Learning with Applications,2026,23:100796.
- [22] HUANG H H. Democratizing LLM efficiency: from hyper-scale optimizations to universal deployability[J/OL]. ArXiv, 2025, abs/2511.20662. <https://api.semanticscholar.org/CorpusID:283261746>.
- [23] SAPKOTA R, ROUMELIOTIS K I, KARKEE M. AI Agents vs. Agentic AI: A Conceptual taxonomy, applications and challenges[J]. Information Fusion, 2026, 126(B): 103599.
- [24] 刘阳, 柏永杰, 林惊. 面向人机物高效融合与协作的具身智能技术体系 [J]. 机器人, 2025, 47 (4): 560-580.
- [25] SimilarWeb. SimilarWeb: DeepSeek 周访问量激增 614%, 峰值访问量达 4900 万次[EB/OL]. 2025. <https://www.similarweb.com/zh-cn/blog/deepseek-traffic-surge/>.
- [26] 中国第一汽车集团有限公司. 基于 5G-V2X 的 L4 级智能商用车车路协同系统 [EB/OL]. 2024. <http://wap.sasac.gov.cn/n4470048/n29955503/n30143035/n30143047/n30143160/c30163453.html>.

## [作者简介]

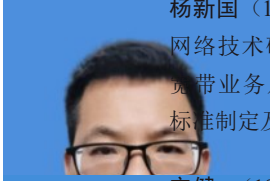


俊 (1988), 女, 中国移动研究院基础网络技术研究所研究员, 主要从事 IP 承载

网及智算 IP 网络相关领域技术研究、标准制定和网络推进；



颜宇 (1978)，女，中国移动研究院基础网络技术研究所项目经理，主要从事 IP 城域网、骨干网领域相关技术研究、标准制定及现网推进工作；



杨新国 (1981)，男，中国移动研究院基础网络技术研究所项目经理，主要从事家庭宽带业务及 IP 城域网领域相关技术研究、标准制定及现网推进工作；



宋健 (1983)，男，中国移动研究院基础网络技术研究所项目经理，主要从事智算 IP 网络新

架构新业务的洞察分析和技术规划等工作；



姜文颖 (1977-)，女，中国移动研究院基础网络技术研究所承载网室技术经理，主要从事下一代互联网、IPv6+基础协议、AI 广域网、超高速以太网等；



程伟强 (1980)，男，中国移动研究院基础网络技术研究所副所长，教授级高级工程师，长期从事网络领域的技术研究和标准制定工作。在 ITU-T、IETF、ONF 等国际标准化组织中主导发布 20 余项国际标准，任 BBF SPAC 委员会副主席、IETF SRv6 头压缩设计组主席，曾任 ONF 电信级 SDN 工作组主席。