Network Innovation and Development Alliance
全球固定网络创新联盟

面向智能体通信的城域云网演进架构及关键技术

—

Network Innovation and Development Alliance

November, 2026

目　录

第一章：智能体通信发展态势 3

1.1 为什么需要智能体通信 3

1.2 产业发展趋势与政策牵引 3

1.3 整体演进趋势 4

1.4 技术发展路线 5

第二章：场景分析 8

2.1 ToH 场景 8

2.2 ToB 场景 9

2.2 ToC 场景 11

第三章：面向智能体通信的城域云网演进架构及关键技术 13

3.1 面向智能体通信的城域网络演进要求 13

3.2 面向智能体通信的城域云网演进架构 16

3.3 面向智能体通信的城域云网关键技术 19

第四章：典型部署模式 24

4.1 基于 vBRAS 构建家庭智能体网关 24

4.2 基于融合边缘打造端云协同智能体网关 25

4.3 基于企业网关升级园区智能体网关 26

4.4 基于云内资源池构建智能体网关 27

第五章：总结与展望 29

第一章：智能体通信发展态势

1.1 为什么需要智能体通信

随着人工智能技术的发展，具备感知、决策与执行能力的智能体（AI Agent）正逐步成为网络中的重要参与主体。在此背景下，智能体互联网（Internet of Agent, IoA）作为一种面向智能体的新型网络范式应运而生。智能体互联网通过统一的通信机制、身份体系与能力发现机制，将分布在终端、边缘云及中心云的各类智能体进行互联，实现智能体之间的安全通信、能力调用与任务协同。

在智能体互联网中，智能体不仅是数据的生产与消费主体，还能够通过网络实现跨域、跨平台的动态协作与资源调度，实现面向任务目标的分布式自主协作。为支撑大规模智能体的并发运行与实时协作，智能体互联网需实现分布式、低时延和高可靠的智能体通信，为智慧家庭、智能制造、智慧交通等各类新型智能体应用提供基础支撑。

1.2 产业发展趋势与政策牵引

伴随智能体生态的逐渐繁荣，具备感知、理解、规划、执行能力的智能体，正逐渐成为人工智能技术价值落地的核心载体。智能体互联网作为面向智能体的新型网络体系，成为下一代互联网演进的重要方向，也为 ICT 产业开辟了全新增长曲线。Gartner 指出，到 2026 年底，40%的企业应用将内置任务型智能体，较去年不足 5%的占比实现跨越式增长。AI 手机、AI PC、自动驾驶汽车、AI 眼镜、人形机器人等智能硬件出货量逐年攀升，据 Canalys、中信证券等机构统计预测，2024-2030 年智能体互联网核心智能硬件市场将实现高速增长，2030 年整体市场空间超 45 万亿元。

我国政府高度重视智能体互联网产业与技术发展，出台了一系列政策文件引导其健康发展。2025 年 8 月，国务院印发《关于深入实施“人工智能+”行动的意见》，指出“到 2030 年，我国人工智能全面赋能高质量发展，新一代智能终端、智能体等应用普及率超 90%，智能经济成为我国经济发展的重要增长极，推动技术普惠和成果共享”；2026 年 1 月，工信部、网信办、发改委等八部门印

发《“人工智能+制造”专项行动实施意见》，提出了“基于国产智能体互联协议，研发高性能智能通信中间件，实现软件与大模型的高效协同及多源数据统一分析”的要求。

1.3 整体演进趋势

智能体互联网的崛起，其深层发展动力源于 AI 能力爆炸式增长，使得 AI 驱动的软件拥有自主性和互相协同的能力。具体而言，智能体通信方式呈现以下几种关键变化：

①互联主体由人向智能体转变

传统互联网中的连接通常以人为中心，由人发起请求、确认操作、配置规则、承担最终责任。因此，其系统设计围绕“如何让人用得方便”展开，前提条件是人员在线，否则会造成流程中断、服务停摆。伴随着人工智能技术的快速发展，智能体成为独立注册、自主发现、可信通信、持续运行的软件实体，它拥有身份标识、服务能力清单、状态心跳和策略边界，逐渐成为与人类同等重要的互联主体。

②交互方式向多模态融合发展

随着大模型能力的跨越式提升，智能体间的协同不再局限于传统的结构化数据或单一文本协议，而是向包含视觉、语音及各类传感器数据的多模态融合方向演进。从能力的角度，智能体能够同时处理来自对端和环境的各类媒体信号以及复杂的逻辑指令，从而在工业自动化、具身智能等高精度场景中完成更复杂的决策。从交互协议的角度，为了支撑海量异构智能体的高效协同，智能体之间的通信正逐步支持多模态数据的直接传递与对齐，通过统一的表征空间，不同模态的信息得以在智能体之间无损互操作，不仅提升了智能体对现实世界复杂任务的适应性，更为构建高度智能化、自主化的智能体互联网络奠定了坚实的交互基础。

③运转方式向高效语义协同发展

传统业务系统运转依赖人工预设规则与周期性人工干预：开发者编写固定的逻辑处理常见场景，运维人员设置阈值告警应对异常，业务人员每日巡检日志排查问题。这类系统的运转行为僵化，面对新组合、新路径、新冲突时普遍失能，需人工介入重启流程或重写策略。而在智能体互联网中，智能体自身具备自主能

力发现、自主工作流编排与自适应冲突解决能力，并且智能体之间能够基于共同目标，语义协商执行协议、动态调整行为策略并持续对齐协同状态。

1.4 技术发展路线

当前智能体通信呈现多路径并行演进格局：

方向一：为快速抢占市场、构建技术护城河，主流云服务商普遍选择构建私有协议实现自有云内的智能体通信。在此背景下，谷歌主导开源 A2A（Agent2Agent）协议，旨在实现不同供应商或基于不同架构的智能体间的安全通信协作，被 Google Vertex AI、Azure AI Foundry、AWS Bedrock Agent Orchestrator 等平台采用。具体而言，A2A 协议定义了智能体间通信的通用语言与交互模型，使不同框架、不同语言、不同厂商构建的智能体能够相互发现、协商协作模式并共同完成用户目标。其采用三层架构设计：第一层为规范数据模型，定义任务、消息、智能体卡片等核心对象；第二层为抽象操作，规定发送消息、流式传输、获取任务、取消任务等必须支持的能力；第三层为协议绑定，提供 JSON-RPC、gRPC 与 HTTP/REST 三种具体实现方式，确保语义一致性与跨绑定互操作。此外，A2A 协议支持同步请求响应、实时流式更新与异步推送通知三种交互模式，适配复杂任务分解、人机协同与长周期执行等多种场景。

方向二：对于网络服务提供商而言，其核心目标是构建支撑海量异构 Agent 跨域互联下的可信通信基础设施，以实现端到端身份可验与行为可溯的智能体接入认证、高效智能体跨域寻址、确定性低时延智能体通信。为此，业界提出 Agntcy、智能体网关等多套方案，致力于构建分层的智能体通信架构。以 Agntcy 方案为例，该方案采用三层架构设计：第一层为身份层，遵循 W3C Decentralized Identifiers（DID）与 Verifiable Credentials（VC）标准，为每个智能体颁发可验证数字凭证，同时支持 Okta、Entra ID、Auth0 等主流身份提供商，也兼容 A2A AgentCard；第二层为发现层，智能体依据 Open Agentic Schema Framework（OASF）发布结构化能力记录，再由智能体目录服务采用分布式哈希表（Distributed Hash Table, DHT）实现技能驱动的全局寻址，并生成内容标识符作为不可篡改的全局唯一地址；第三层为传输层，即安全低时延消息协议（Secure Low-Latency Interactive Messaging, SLIM），其核心是可独立部署的轻量级网络功能实体 SLIM

Node，承担消息中继、Message 端到端加密、多跳路径建立与会话管理等职责，形成覆盖终端、边缘与云的统一通信平面。

以上两种技术方向由于面向场景不同而呈现能力边界分化：面向云侧智能体通信场景设计的 A2A/ANP 等协议，是应用层点到点互联的架构，应用在大规模的智能体互通场景时，存在可扩展性差、不支持群组通信、智能体交互缺乏管控等问题；采用分层设计的 Agntcy、智能体网关等方案，则综合考虑应用层与网络基础设施层的一体化协同，通过引入新的网络节点（SLIM Node/Agent Gateway），从而为应用层 A2A 协议提供群组通信、交互管控等能力。两种技术方案在寻址发现、身份认证、通信方式的对比如下：

表 1：两种典型方案对比

	A2A	Agntcy
发现寻址	基于结构化文件（如 AgentCard），描述 Agent 能力、要求、签名等信息。调用方通过域内搜索引擎或爬虫获取该文件，解析能力描述与 URL 地址后即可发起调用	Agent 主动向分布式目录节点注册，提交能力描述、身份标识等信息，目录节点为其生成唯一标识（Agent ID），并通过 DHT、ANS（Agent Name Service）等技术实现跨域 Agent 能力共享。调用方基于语义向目录节点查询 Agent，由目录节点搜索适用 Agent
身份认证	Agent 描述文件中包含数字签名、认证要求等身份信息，并包含 OAuth 2.0、mTLS 等多种认证方式。调用方据此自主完成 OAuth 2.0 等认证流程向 Agent 发起请求；Agent 自身是否安全可基于签名和日志审计等机制监督评判	Agent 需向具备 Issuer 资质的、可验证的权威实体（如合规审计机构、云平台合规中心、行业联盟等）申请 VC，在注册时需向分布式目录节点提交 VC 以供验证 Agent 身份。调用方在发起请求前，需获取目标 Agent 的 VC 执行实时校验
通信方式	采用直连方式，调用方解析 Agent 描述文件后，直接向其声明的 URL 发起 JSON-RPC、gRPC、HTTP 等	跨域通信基于网关实现（实现 SLIM Node 等中继节点功能），当调用方请求到达网关后，由网关

	类型的请求，无中间协议转换或策略干预；Agent 自主处理请求并返回响应	进行协议转换、路径选择、安全增强等操作，再转发至目标 Agent 所在网关, 由该网关寻址转发至目标 Agent

第二章：场景分析

2.1 ToH 场景

2.1.1 场景概述

（1）家庭终端智能体协同场景

该场景指家庭内部部署的多个具备感知、计算或执行能力的异构终端智能体，如智能管家、智能摄像头、服务机器人、智能音箱等，通过安全高效的域内协作，使数据在家庭内汇聚和处理，并在本地生成相应决策或控制动作，从而在保障各类终端智能体的低时延响应与高实时性控制的同时，降低敏感信息外传风险，保障家庭数据隐私。其中，智能管家智能体作为家庭智能体网络的协调节点，承担着任务理解、意图解析、能力编排等职能，一方面需要接收用户指令理解用户需求，另一方面按需调用其他家庭智能体完成相应服务。例如，智能管家智能体调用摄像头进行环境感知、调用机器人执行巡检或交互任务、调用智能音箱播放信息等。

（2）家庭终端智能体与云端/边缘智能体协同场景

该场景指家庭内部终端的算力、存储能力或服务能力有限，难以独立支撑高复杂度 AI 任务，通过与云端或边缘智能体进行协同处理，有效提升复杂任务处理准确率。在该场景中，终端智能体采集的环境感知数据或任务请求可以脱敏后上传至云端或边缘，由具备更强推理能力的智能体完成图像识别、语义理解或多模态内容生成等计算密集型服务。云端处理完成后，将结果返回至智能管家智能体，由其结合家庭内部状态进行进一步决策，并协调终端智能体执行相应操作。以智慧康养场景为例，摄像头检测到老人疑似跌倒时，将去标识化信息上传至云端/边缘，云端/边缘智能体结合历史行为模式完成高置信度跌倒判定并生成处置建议，告知智能管家智能体进行安抚，并向家属推送摘要通知，全过程原始视频数据始终留存于本地，仅特征与决策结果跨域流动。

2.1.2 场景核心需求

ToH 场景由于家庭环境的高隐私性、用户低专业性、终端算力局限性及网络条件复杂性，对智能体安全可信接入、本地协作和低时延通信等提出了独特要求。在此背景下，ToH 场景的核心需求主要包括以下几个方面：

- **异构智能体寻址与通信**：家庭内部存在多种异构智能体，它们之间需要实现能力同步及跨协议互操作，确保任务调度和状态同步的实时可靠性。
- **高隐私保护**：家庭成员的个人信息及隐私喜好等长期记忆的 Agent 数据库等高敏感数据必须尽可能在本地处理和存储，避免隐私泄露。
- **安全可信监管**：家庭内部所有智能体需具备可信数字身份与可控安全权限，交互协作的通信及内容按需记录，实现防攻击、防窃取和可监管。
- **通信效率与实时性**：ToH 场景的需求常具有实时性，例如，家庭监控摄像头的异常识别与动作响应需在 500ms~1s 内完成，需要低时延通信保障。

2.2 ToB 场景

2.2.1 场景概述

（1）终端智能体跨园区协同场景

该场景指部署在不同物理园区、地理分散但业务关联的终端智能体，通过可信通道直接交互状态、能力与策略，无需依赖中心平台调度，即可完成跨域感知、联合决策与协同执行，从而提升整体系统的响应速度、容错能力和资源利用效率。以电力调度场景为例，多个地市级变电站各自运行的负荷预测智能体、设备状态智能体与储能调度智能体，通过安全可信通道实时交换本地实时负荷曲线、主变健康度评分及电池可用量；当 A 市电力负荷陡增、备用容量跌破阈值时，其负荷预测智能体主动向 B 市、C 市同类智能体广播协同请求，各站智能体基于本地约束条件（如设备检修窗口、上级调度指令锁定状态）自主响应，并联合生成区域级错峰方案，全过程仅确认过程需中心平台参与，实现分布式智能、高韧性电力保供。

（2）企业智能体端云协同场景

该场景指部署在企业园区内的智能体与云端智能体，通过分层任务分配与轻量化信息交互，在保障低时延控制的同时，充分调用云端算力与全局知识，提升单点执行精度、系统运行效率与持续进化能力。以智能工厂生产调度为例，某汽车零部件产线上的设备状态智能体实时采集振动频谱与温度时序数据，并仅压缩上传异常特征向量至云端；云端工艺优化智能体结合全集团近 3 年同类设备故障图谱、材料批次参数等信息，完成根因定位并生成参数微调建议；该建议以策略包形式下发至园区设备智能体，由其在下一个工单间隙自主测试并反馈效果数据；同时，云端知识中枢智能体将本次闭环经验结构化为新规则，同步注入各工厂同类产线的质检智能体与预测性维护智能体，实现“一厂诊断、百厂受益”。

2.2.2 场景核心需求

ToB 场景因组织目标的强约束性、业务数据的强主权性、决策流程的强角色性及智能体的跨域异构性特征，对智能体间的通信机制提出了专属要求，核心通信需求主要包括以下几个方面：

- **跨域语义寻址**：智能体分散部署于企业内网、边缘设备、私有云、行业云及合作伙伴系统，通信发起方通常仅知晓目标所需能力与策略条件，要求通信机制能基于能力语义与多维策略自动定位并连接合规可用的智能体。
- **端到端安全通信**：企业智能体通信场景中，原始业务数据严禁跨域明文流动，要求数据交互在发送方授权范围、接收方策略边界与第三方审计框架内完成，原始数据始终留存在主权方本地系统。
- **意图协商通信**：ToB 业务天然具备长链条、多角色、强规则特征，单次请求无法承载完整意图与约束条件，要求通信机制支持多轮、带状态、可回溯的意图交换，使智能体能在动态协商中逐步对齐前提、承诺、交付证据。
- **确定性 SLA 通信保障**：ToB 关键业务通信对时延稳定性、报文顺序性与交付成功率存在硬性契约要求，要求通信机制保障确定性时延、零抖动、高交付成功率，使每次交互均满足预设 SLA 并可验证履约结果。

2.2 ToC 场景

2.2.1 场景概述

（1）个人终端智能体端云协同场景

该场景指部署在用户手机、电脑等端侧设备的智能体与云端服务智能体，通过“端侧敏感数据本地处理 + 云端复杂任务深度推理”的分工模式，实现端侧快速响应与云端全量服务的协同增益。以个性化 AI 助手场景为例，当用户提出“规划下周末出行计划”的需求，端侧助手智能体完成用户需求的语义解析与任务拆解，在本地闭环处理用户日程、出行偏好、人员信息等敏感隐私数据，仅将脱敏后的需求特征上传至云端；云端同步调度多领域服务智能体，并行完成公域信息检索、资源核验与方案整合；最终由终端智能体完成方案的个性化适配，输出全流程一体化的可执行出行方案。

（2）终端多智能体协同场景

该场景指部署在手机、平板等各类移动终端的智能体，无需依赖云端平台全程调度管控，可在弱网、无公网环境下自主完成身份互认、任务协同与联合执行，构建低时延、高隐私的分布式个人智能体协同网络，同时支持不同用户终端智能体的可信点对点联动。以高铁出行场景为例，用户工作机、私人机两台移动终端中分别部署智能体，在列车弱网环境下无需云端中转，即可通过局域网完成点对点可信身份核验。用户发起差旅报销需求时，两台终端智能体分别在本地处理工作纪要、行程票据等分区数据，点对点同步进度并联合生成报销草稿，核心数据全程仅在终端间流动，兼顾弱网环境下的业务可用性与数据隐私安全。

2.3.2 场景核心需求

ToC 场景因用户需求的个性化、数据的私域化、交互的多模态化等特征，对智能体的协同调度、隐私保护、精准响应和低时延通信等提出了专属要求，核心需求主要包括以下几个方面：

- **协同编排与能力调度**：用户侧存在多类型、分布式部署的异构智能体，涵盖本地端侧与云端部署形态，需实现智能体能力的统一注册、按需发现与高效

编排。

- **私域数据安全与隐私保护**：用户拥有大量个人日程、本地文件、偏好画像等高敏感私域数据，此类数据需在本地智能体间完成处理与交互，严禁无授权向公域传输。
- **个性化需求精准响应**：用户需求呈现多元化、复合型特征，且存在鲜明的个人语境与使用习惯，需融合用户历史偏好、实时场景、行为反馈等多维度信息，实现需求的精准解析与个性化结果输出。
- **低时延与高可靠通信**：ToC 场景的交互多为实时性需求，如 AR 眼镜实时翻译、车载 AI 即时响应等，需保障跨智能体通信的低时延与高可靠性。

第三章：面向智能体通信的城域云网演进架构及关键技术

3.1 面向智能体通信的城域网络演进要求

智能体互联网将驱动网络流量模型、业务形态和资源部署模式发生深刻变革。作为连接用户与算力的核心基础设施，城域网络需要从尽力而为的管道向“面向智能体的智能承载底座”演进，主要体现在承载平面、流量模型以及管控精度等多个维度。

3.1.1 业务承载

（1）运营商自营 IoA 业务承载要求

作为基础网络提供商，运营商可以构建并运营自有的智能体互联网服务，面向家庭、政企与个人提供高效、安全、灵活的智能体互联解决方案。综合 ToH、ToB、ToC 场景的核心需求分析，可归纳为以下几点：

- **身份**：智能体在接入网络时，应当被赋予一个全局唯一、与物理基础设施无关的智能体身份标识。同时，智能体身份管理可与运营商已有用户管理系统配合，通过建立用户 ID 与智能体 ID（Agent ID）的可信映射关系，支持用户身份策略向智能体自动继承，确保人机身份统一、策略一致、责任可溯。
- **接入与鉴权**：支持海量智能体的准入控制、认证与鉴权，确保接入实体的合法性与身份可信。由于智能体具备调用外部工具或执行实体操作的能力，需构建全链路“零信任”体系，使身份认证、权限授信贯穿于通信的每一个环节，确保只有经过授权的智能体才能在限定的上下文环境中进行有限交互。
- **注册与发现**：注册时需明确异构智能体的通信方式、能力语义、运行约束及安全策略等信息，确保协同发起方能够准确理解其可提供服务的边界、前提条件与调用风险。此外，还需具备高效的、分布式的智能体能力目录，以支持基于语义的智能体能力注册与跨域发现机制。
- **通信与协作**：支持面向复杂的多智能体协作编排与动态群组通信，并实现面向智能体通信的高效路由与灵活组网。同时，还需支持异构协议转换与端到端加密通信，同时按需提供差异化的通信保障，构建安全、灵活、高效的智

能体通信底座。

（2）端侧业务承载要求

端侧通用智能体广泛部署于手机、个人电脑等个人使用的终端设备，或者无人机、车载单元、机器人等具身智能体上，其承载需求总体上与运营商自营 IoA 业务一致，均要求城域网络提供身份可信、接入可控、可发现可协同、安全高效的通信环境。但端侧智能体受限于算力、功耗及接入方式多样性，存在以下差异化要求：

- **身份**：同样需具备全局唯一的身份标识，但需兼容更广泛的生态，支持与第三方身份体系（如设备厂商 ID、云服务商 ID 等）的互认，以支撑跨厂商、跨平台的通用智能体可信互通。
- **接入与鉴权**：需支持海量端侧智能体的并发接入与轻量级、低时延认证，支持证书、OAuth 2.0、DID/VC 等多种鉴权方式，并对移动性强的智能体（如车载 Agent）提供接入点切换时的快速重认证与上下文迁移。
- **注册与发现**：需充分考虑端侧智能体频繁上下线的动态特性，支持注册信息的自动更新与过期清理，并在边缘位置提供就近的注册与发现服务。同时支持基于局域网的本地发现机制，使端侧智能体在不依赖广域网时仍能完成本地协同。

（3）公共 IoA 业务承载及管控要求

在跨厂商、跨生态的公共环境中，城域网络作为中立且广泛覆盖的基础设施，其核心任务是消除异构智能体间的通信屏障，并提供企业级的 QoS（Quality of Service）保障。对于跨企业分支或跨云平台的智能体协作，网络应能提供类似专线的确定性管道，确保关键任务数据不被普通互联网背景流量所影响。

更进一步，网络应具备与应用的深度联动能力。城域云网需要能够感知智能体应用的需求，例如，通过感知 Agent ID、类型、能力等关键信息，并将其自动、实时地映射为网络层的转发策略、QoS 策略和安全策略，实现应用感知和端网协同，为上层智能体应用提供易用且有效的网络服务。

3.1.2 新型流量模型

智能体通信将改变传统互联网以下行流量为主的特征。例如，在传统的家庭

宽带（ToH）场景中，用户行为主要表现为视频流观看或网页浏览，网络设计侧重于优化下行带宽。然而，随着智能家居智能体、具身智能设备以及多模态交互的普及，智能体需要与云端或边缘节点高频交换结构化事件、实时状态快照及闭环协作反馈，使得上行流量呈现爆发式增长；在企业（ToB）场景中，例如 AI 编程等应用，也会涉及海量工程文件的上载。这些应用都会让传统的非对称带宽分配策略面临挑战。

除了上行流量增长，在未来大规模智能体协作中，云间、端端跨域协同还将进一步催生东西向流量增长。多个区域的智能体为完成同一任务，可能在短时间内发起高并发的东西向通信请求，并且可能会维持长连接。这些都是当前城域网所未承载过的流量模式。

3.1.3 精细化 Token 经营

大模型服务正成为新的数字消费主体，这对运营商的计费和网络能力提出了新要求。从用户视角出发，如果运营商能提供“一站式”入口，统一访问多个大模型厂商的服务，并实现统一的 Token 计费，将解决用户在多个模型厂商间独立开户、购买套餐带来的碎片化及浪费的痛点。运营商可借此从管道提供商向“Token 分销商”转换，显著提升流量经营的附加价值。更进一步，运营商还可以充分发挥广泛分布的边缘计算资源，为大模型/智能体厂商提供靠近企业/消费者用户的就近推理服务，从“Token 分销”升级为“Token 生产”，显著提升运营商在生态中的价值。

为实现精细化 Token 运营，城域网应能识别智能体推理相关的流量，并感知其所属的模型、优先级、安全等级等属性，并为不同类型的推理流量提供差异化的 QoS 保障。例如，网络应优先保证首词延迟（TTFT, Time-to-First-Token），并在 Token 生成过程中维持稳定的吞吐率，将运营商的传统强项在智能体新兴业务中再次实现价值变现。对于 Token 边缘生产，运营商可结合网络路由等技术，将推理相关的流量引流到合适的边缘进行就近处理，提升用户体验。

3.2 面向智能体通信的城域云网演进架构

3.2.1 城域云网构建思路

云业务的蓬勃发展持续驱动着云和网在城域范围内的融合发展。在这一背景下，业界逐步形成共识：城域网络需要打破传统以业务场景为导向的建网模式，从网络架构重构、承载技术创新和运营维护统一等方面着手，实现与云/算力深度融合，构建以“云”为中心的城域云网。

- 架构层面：借鉴云内网络的设计思想，在城域内引入基于 Spine-Leaf 的“积木式”弹性可扩展架构。这一架构打破了传统城域网的三层或多层层级结构，实现弹性组网和业务灵活接入，以及广域东西向与南北向流量的无阻塞快速转发。
- 技术层面：基于 SRv6/EVPN 构建可编程的端到端云网平面，为固、移、云、算多业务融合承载奠定基础。基于 SRv6 实现业务级的流量工程，满足不同业务对时延、带宽和可靠性的差异化需求；基于 EVPN 实现二层和三层业务的统一承载，简化传统网络中 VLAN、VPLS 等多套控制协议并存的复杂局面。
- 运营层面：基于统一云网运营系统，打破传统网络和云平台各自独立运维的壁垒，实现云网资源的统一管理、统一编排和统一运维。一方面需将网络资源（带宽、端口、隧道等）与云资源（计算、存储、容器等）纳入统一资源池进行建模和管理，形成全局的资源视图；一方面需基于标准化 API 接口、数字地图等方式，实现云网业务的一体化编排和端到端性能监测。

目前，从中国电信到瑞士电信等全球领先运营商，都已将上述架构与技术路线纳入其网络演进的核心战略，并开展了相应的试点与商用部署，推动城域网络向“以云为中心”的城域云网演进。

3.2.2 城域云网面向智能体通信的关键变化点

面向智能体通信带来的新需求，城域云网将在以下几个关键方面进一步演进：

- 容量与拓扑：虽然基础物理架构保持不变，但由于智能体带来的上行流量和

东西向流量激增，网络各层级的带宽配比需进行定量调整。接入层到汇聚层的收敛比将缩小，甚至在算力密集的区域需要采用非阻塞的 Full-mesh 连接方式。

- 引入智能体网关（Agent Gateway, AGW）：AGW 作为一种新型网元，是城域网络演进中最重要的变化。AGW 作为智能体的“通信汇聚点”和“安全锚点”，负责执行身份认证、寻址、访问控制等关键任务，可以部署在家庭 CPE、城域网边缘（如 BNG/PE）或云边界。
- 构建智能体通信新平面：基于 AGW，在现有物理网络之上构建一个独立的新平面。该平面基于新型承载底座，具备面向智能体通信的独立控制面和独立转发面，为大规模智能体互联提供高效、安全、隔离的承载环境。
- 面向智能体通信的 NaaS（Network as a Service）服务和接口：通过 NaaS 的发展与推进，网络将面向智能体应用层开放实时组网、确定性路径请求、流量加密等接口。智能体可以像应用程序调用 Socket 接口一样，方便地调用各类网络服务，实现资源随业务需求的即时建立与拆除。

3.2.3 面向智能体通信的城域云网目标架构

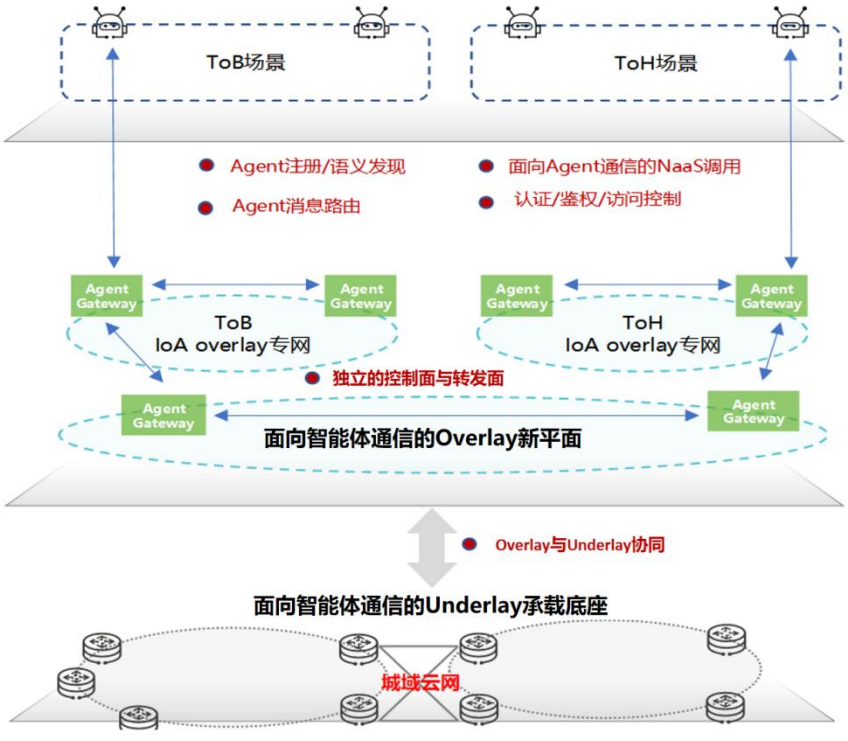


图 1 面向智能体通信的城域云网目标架构

如上图所示，面向智能体通信的城域云网将在当前基础设施之上演化出一张面向智能体通信的新平面，以此专网作为承载智能体通信的主体。智能体通信平面由新引入的智能体网关组成，通过 overlay 的方式在基础设施之上互联，拥有自己的控制面和数据面。

同时，underlay 网络需要演进为可被感知、可被调度、可被承诺的智能承载基座，支撑智能体通信平面的按需构建与动态协同。它不再仅是静态转发管道，而是具备语义理解能力的服务化基础设施，能主动适配智能体通信在时延、可靠性、策略合规等方面的差异化需求，实现与 overlay 平面的双向使能。

3.2.4 面向智能体的城域云网平滑演进路线

面向智能体通信的城域云网演进，并非对现有网络进行重构，而是在当前已具备优势的城域云网基础上，围绕关键变化点，按照“带宽先行、智能体网关增量部署、新平面逐步构建、NaaS 能力开放”的内容和节奏，实现从传统云网融合基础设施向智能体原生承载底座的平滑过渡。

- **带宽向 400G/800G 升级**：智能体互联网的兴起将深刻改变流量模型：ToH 场景中多模态交互与实时状态上报将推动上行流量爆发式增长，ToB 场景中跨园区协同与端云协同将催生海量东西向高并发通信需求。为承载智能体通信带来的高并发、大带宽业务流，城域骨干层需向 400G/800G 平滑演进，接入与汇聚层则需适度降低收敛比，在算力密集区域可按需采用非阻塞 Full-mesh 互联，确保端到端无阻塞转发能力。
- **智能体网关增量部署**：智能体网关是本次演进中最重要的新型网元，其部署遵循业务需求驱动、分场景覆盖的原则，优先在需求最迫切的位置落地。面向家庭场景（ToH）：家庭是智能体接入量最大、终端异构性最强的场景。可优先将宽带接入网中的 BNG/vBRAS 设备升级为具备智能体网关功能的节点。面向企业场景（ToB）：企业场景对智能体通信的确定性、安全隔离和跨域协同要求更高。初期可按需在园区出口或企业云接入点部署独立的 AGW，优先满足各行业对实时控制、数据主权等关键特性的需求，实现园区内智能体的本地闭环与跨园区协同。中期随着企业智能体规模扩大，可在园区侧、接入侧、云侧分层部署 AGW，形成“园区自治+跨域协同”的分布

式智能体网络。

- **新平面构建**：在 AGW 逐步部署的基础上，依托城域云网物理基础设施，构建独立的智能体通信新平面，为大规模智能体互联提供注册/发现/消息路由等服务；并充分结合 Underlay 网络已具备的 SRv6、网络切片等能力，在不改变物理网络拓扑的前提下，实现逻辑上的独立性与业务上的差异化保障。在实现层面，可结合 APN6（应用感知网络）技术，将智能体应用的业务属性 (如 Agent ID、任务类型、QoS 等级) 封装在 IPv6 报文中，传递至 Underlay 网络，实现 overlay-underlay 无缝协调。
- **NaaS 开放**：为使智能体应用能够按需、实时地获取网络服务，城域云网需面向新平面开放 NaaS 能力。将 Underlay 网络的带宽、切片、时延、路径等资源抽象为可调用的网络服务，智能体可通过与智能体网关的信令交互，声明具体的服务需求(如“在智能体 A 与 B 之间建立低时延通道, 时延<50ms”)，自动触发网络资源预留与策略下发。这种开放模式使智能体能够像调用计算和存储资源一样方便地调用网络服务，为精细化 Token 经营、动态虚拟专网创建等场景提供能力支撑，真正实现“网络即服务”。

3.3 面向智能体通信的城域云网关键技术

3.3.1 智能体网关技术、架构与实现方式

智能体网关不仅是一个简单的路由器，它是一个集成了计算、存储与通信能力的复合型网元，其技术要求如下：

(1) 原生功能集

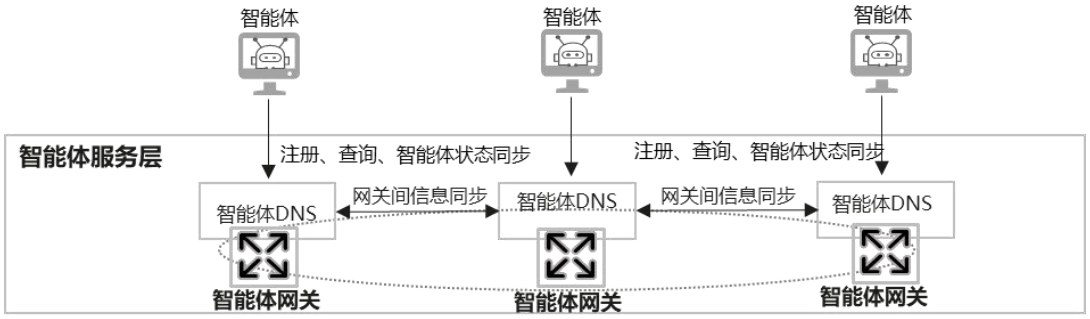


图 2 智能体能力注册与发现

- **智能体能力注册发布**

为实现智能体的统一描述与互通，需要定义标准化的智能体描述模型，包含智能体 ID、名称、位置及技能标识符等关键信息，实现对智能体的规范化描述与唯一标识，为后续的发现与交互提供基础。

对于需要接入城域网的智能体，需要通过智能体网关进行注册。网关在接收智能体注册请求后，对智能体身份进行认证。认证通过后，为其分配一个通用且唯一的标识，即智能体 ID，并将该 ID 返回给智能体。注册完成后，智能体即可被网络中的其他实体感知与使用。同时，智能体网关基于智能体身份/权限进行访问控制，并根据具体的管理需求记录智能体行为的关键元数据（例如访问目标、数据量等），实现智能体行为的可观测。

- **智能体信息同步**

在多网关环境下，为支持跨域智能体协作，需要在不同网关之间共享智能体相关信息，包括状态信息（如上线、下线）以及能力描述信息，保障多个网关间信息更新的时效性与一致性。为避免信息泛洪，同时实现权限控制与安全隔离，智能体网关间的信息同步需要基于策略进行限制，例如按照管理域、信任关系或智能体业务范围进行选择性同步。

- **智能体寻址**

当智能体需要调用其他智能体能力时，可根据任务需求生成寻址请求，并发送至所属智能体网关。网关基于已有的智能体信息，在本地或跨网关范围内进行匹配，完成目标智能体的定位。寻址过程既可以基于智能体 ID 进行精确定位，也可以基于任务语义或能力描述进行匹配，从而支持面向能力的动态发现。

完成目标智能体定位后，网关或智能体可结合业务需求对目标智能体进行身份认证和可信性校验（如证书验证、访问凭证、信任关系校验等），以保障跨域交互的安全性。在通过认证后，双方即可建立通信关系并开展后续业务交互。

- **智能体间路由**

智能体之间消息的路由具体可分为以下几种情况：

- ✓ 点到点路由：可复用基础设施层的 IPv4/IPv6 路由，无需进行特殊处理，但需要智能体来维护通信对端的 IP 地址和可达性；或者，由智能体通信平面根据智能体的身份标识等元数据进行选路，从而让智能体专注在业务层面的处理，免于应对 IP 地址动态更新、移动性管理等网络层的复杂性。

- ✓ 群组通信路由：可复用基础设施层的 IP 组播技术，同样需要智能体支持 IP 组播组建立、加入等功能；或者，由智能体通信平面接管群组通信，以智能体网关作为连接锚点，根据智能体的需求，实时创建通信群组、并实现组内的广播、订阅-发布等通信模式。

- **智能体间通信**

在完成智能体发现与寻址后，智能体之间通过统一的通信机制进行交互，实现任务请求、结果返回以及状态反馈等功能。智能体通信机制以智能体 ID 为核心标识，在消息交互过程中显式携带源智能体 ID 与目的智能体 ID，从而支持跨网关场景下的消息定位、转发控制与安全审计能力。智能体之间通过所属智能体网关进行通信，网关基于智能体标识完成消息转发，从而实现跨域互联互通。

（2）架构设计

- 控制面：AGW 之间的交互协议，用来同步已注册的智能体记录，从而实现分布式的智能体发现与基于 Agent ID/语义的路由机制；此外，通过控制面协议，也可实现智能体访问控制策略的同步。
- 转发面：根据 Agent ID 的转发，形成基于 Agent ID 的路由表，与 Underlay 的 IP 路由表解耦。面向长期演进，需要根据智能体当前的任务/需要的能力等信息进行转发，将“发现-转发”行为融为一体，实现基于语义的路由。
- 管理面：对下挂的智能体提供可观测服务，包括智能活动监测、业务质量感知、计费、告警等功能，并向上层业务平台开放北向接口，实现应用与网络的联动。

（3）实现方式

- 硬件方式：基于 NP（网络处理器）或 ASIC 的专用硬件设备，具备高性能、低延迟、高可靠性的优点，适合在城域网核心位置或对性能要求极高的场景部署。
- 软件方式：基于通用 x86/ARM 服务器或云原生环境部署的虚拟化网元（VNF），具备弹性伸缩、快速迭代、成本灵活的优点，适合在边缘云或业务量初期部署。
- 软硬结合：软件部分（运行于通用 x86/ARM 服务器或云原生环境）负责承载 AGW 的控制面与管理面功能，包括智能体注册与发现、语义解析、安全

认证与访问控制策略执行、行为观测与可观测性数据上报等。这类功能逻辑复杂、迭代频繁，对计算能力要求高但对转发时延不敏感，适合以软件方式弹性部署。硬件部分（运行于路由器/交换机之内），承载 AGW 的高性能控制面/转发面功能，如智能体元数据（AGW 下挂的 Agent ID、能力、IP 地址等信息）同步、基于 Agent ID 的路由表生产与转发等。

3.3.2 高效承载技术

为支撑智能体专网的高效运行，城域云网需实现 Overlay/Underlay 的深度协同。

（1）Overlay/Underlay 协同机制

智能体专网作为 Overlay 平面，需要与作为 Underlay 的城域云网基础设施高效联动。如前章节所述，基础网络通过开放服务接口，将自身的网络拓扑、带宽、时延、切片等能力暴露给上层。基于该接口，智能体专网控制器可以：

- 动态路径规划：根据智能体间通信的 QoS 要求（如低延迟、高可靠），实时调用 Underlay 的 SRv6 能力，为其计算并建立最优的端到端路径。
- 资源按需预留：为关键智能体业务（如跨域协作）动态申请 Underlay 的网络专线/切片资源，实现带宽和时延的确定性保障。

（2）Underlay 强化机制

为支撑智能体通信平面的高效运行，Underlay 城域云网基础设施需在无损承载能力、资源隔离与确定性保障能力、可视化与快速自愈能力方面进行强化，为不同 SLA 需求的智能体业务提供精细化、差异化、可承诺、可感知的服务质量。具体可实施的技术包括：

- 无损承载：面向海量 Agent 算力需求，构建端到端流量感知、快速调度与无损保障能力，全面满足分布式推理对低丢包、低时延、高吞吐的极致需求，实现智算中心、智算边缘、智算终端之间的高性能互联。
- 网络切片：利用 Flex-E、小颗粒切片等技术，在 Underlay 网络上创建多个逻辑隔离的“智能体通信切片”。例如，可以为时延敏感的工业控制类智能体创建独立的硬切片，为普通消费类智能体创建软切片，实现业务间的严格隔离和差异化 SLA 保障。

- IOAM：实现随流业务检测，实时感知 Underlay 网络的故障和拥塞，并迅速调整 Overlay 路由，实现毫秒级的业务自愈。

3.3.3 Token-Aware 智能感知与调度技术

为实现对词元（Token）流量的精细化经营，需要引入 Token-Aware 技术。该技术基于智能体网关的 Agent 信息管理，在网络转发面引入轻量级 DPI 或 AI 推理模块，从而实时识别并标记智能体/大模型流量的关键属性，如：

1) 首 Token 流量识别：优先保障首 Token 相关流量的时延，提升用户体验。
2) 模型归属：识别 Token 所属的大模型（如 DeepSeek、Qwen 等）。
3) 业务类型：区分数据流量是来自普通对话、代码生成，还是图像创作等不同的业务。
4) 服务等级：根据用户订阅的套餐，标记业务流量的优先级（如金牌、银牌、铜牌）。

基于这些标记，网络可以执行差异化的 QoS 策略：为高优先级业务流量预留专用队列和带宽，保障其在拥塞时优先通过；同时，这一机制也为基于 Token 产生/分发的计费新量纲提供了精准的网络层数据支撑，使运营商能够从“流量管道”升级为“智能服务编排者”。

此外，运营商也可以充分发挥其广泛分布的边缘算力资源，通过识别推理类流量及其算力需求（如模型类型、输入规模等），结合城域云网的分布式边缘节点布局，利用网络路由策略将推理请求动态引流至最优边缘推理节点就近处理，实现 Token 在边缘侧的生产与返回，有效降低端到端时延，提升推理业务的用户体验。

第四章：典型部署模式

4.1 基于 vBRAS 构建家庭智能体网关

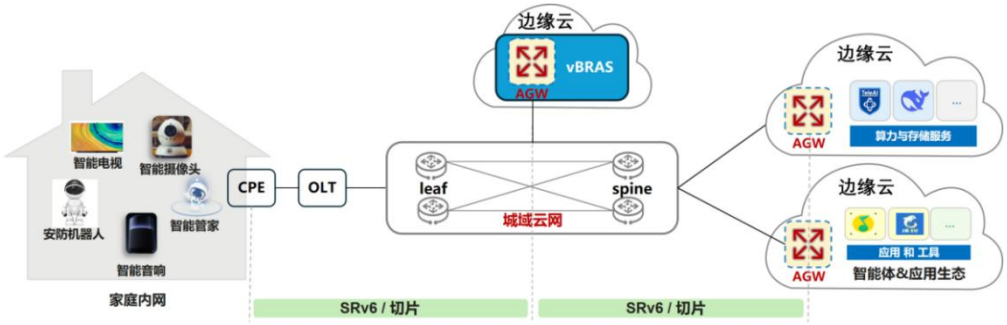


图 3 基于 vBRAS 构建智能体网关

vBRAS（虚拟化宽带远程接入服务器）是城域宽带网络的核心网元，由控制平面（vBRAS-CP）、转发平面（vBRAS-UP）及管理编排系统（MANO）三部分构成。其关键特征在于：采用微服务架构，支持控制与转发分离、软硬解耦；具备海量 PPPoE/IPoE 会话管理能力、细粒度 QoS 策略执行以及与 AAA 系统深度绑定的认证能力。作为家庭宽带用户的接入锚点与会话中枢，vBRAS 天然掌握用户身份、在线状态、IP 地址、带宽授权、终端类型等全维度上下文信息，这使其成为承载智能体网关的理想物理载体：既无需新增硬件节点，又能复用其高可靠会话管理引擎与成熟认证体系，为 AGW 提供低时延、高一致性的身份锚定与策略执行底座。

- **典型场景分析：家庭内终端智能体协同**

家庭域内智能体可依赖本地广播或轻量级爬虫实现相互寻址，但此类方式缺乏统一身份锚点与策略执行能力，极易引发仿冒接入、未授权调用及敏感数据泄露等安全风险。vBRAS（AGW）通过将家庭所有终端智能体的注册、认证与授权深度耦合于用户宽带会话生命周期，在智能体首次接入时即完成其可信数字身份核验与证书签发；并且可基于 vBRAS（AGW）实现智能体通信的严格访问控制。例如，智能管家调用摄像头进行环境感知、服务机器人响应语音指令等协作行为，需在 vBRAS（AGW）的策略监管下完成端到端加密与细粒度权限裁决，原始视频流、设备控制指令等高敏内容始终封闭于家庭本地网络。

4.2 基于融合边缘打造端云协同智能体网关

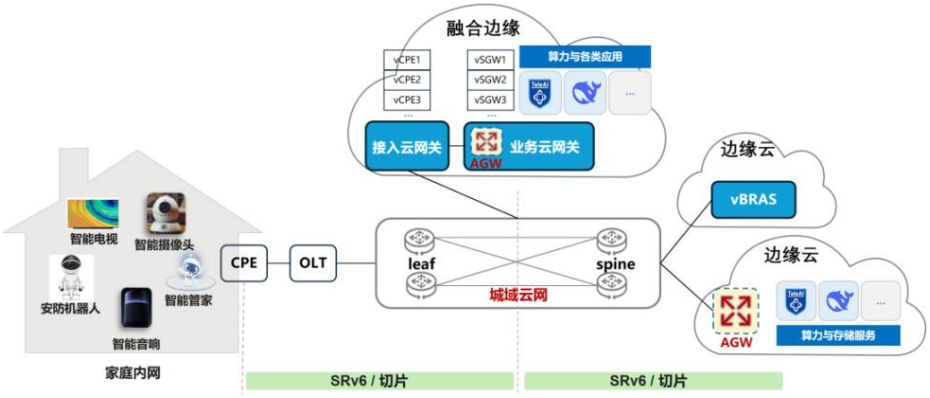


图 4 基于融合边缘构建智能体网关

融合边缘是基于中国电信打造的云网融合新型基础设施，其核心是将传统分散部署在家庭 CPE、小区 BRAS 及中心机房的网络功能重构并下沉至城域边缘云节点。其中，接入云网关、业务云网关均采用转控分离架构设计，以容器化的方式部署。转发面采用 NP 转发模式，提升转发的性能和可靠性，控制面通过云化方式，部署在边缘云内，提升软件升级的灵活性。具体而言，接入云网关融合 vCPE 功能，实现终端接入控制、DHCP Server、PPPoE 拨号、DNS Relay 等功能，作为用户上网、入云业务的分流和转发锚点，将用户流量根据应用分发到边缘云内的业务云网关，上网流量直接送往 UP 访问外网。业务云网关融合 vSGW 功能，实现应用私网 IP 申请、MS 协议代理、双向 NAT 控制面等功能，作为云业务的分发点，将用户访问业务系统的流量进一步分发至对应云应用。

在融合边缘模式中，接入云网关、业务云网关分别作为用户侧接入控制与云侧业务调度的关键角色。接入云网关聚焦家庭智能体可信管理，深度协同 vCPE 端侧信息，实现家庭智能体设备级认证、严格访问控制，确保低时延、高可靠、强合规；业务云网关则聚焦云内智能体管理，记录云内智能体的能力描述、协议适配、服务地址、计费等信息。两者基于统一 Agent ID 与上下文令牌实现双向信息同步与策略协同，并且通过云内的大二层组网保障家庭智能体高体验。

- **典型场景分析：家庭终端智能体与云端智能体协同**

在该场景下，基于融合边缘的部署模式，通过“家庭智能体-接入云网关-业务云网关-云内智能体”的大二层组网，为端云智能体协同提供端到端确定性通信服务；通过“先入云、后入网”的流量调度机制，绕过传统南北向流量迂回路

径，使家庭侧与云侧智能体间的请求响应、状态同步、指令下发等关键通信行为，在更低网络跳数、更少协议转换、更短传输路径下完成，显著提升智能体通信效率。此外，因为接入云网关与业务云网关同属一个城域边缘云节点，智能体能力目录共享无需跨网络同步，天然具备强一致性和毫秒级响应能力。

4.3 基于企业网关升级园区智能体网关

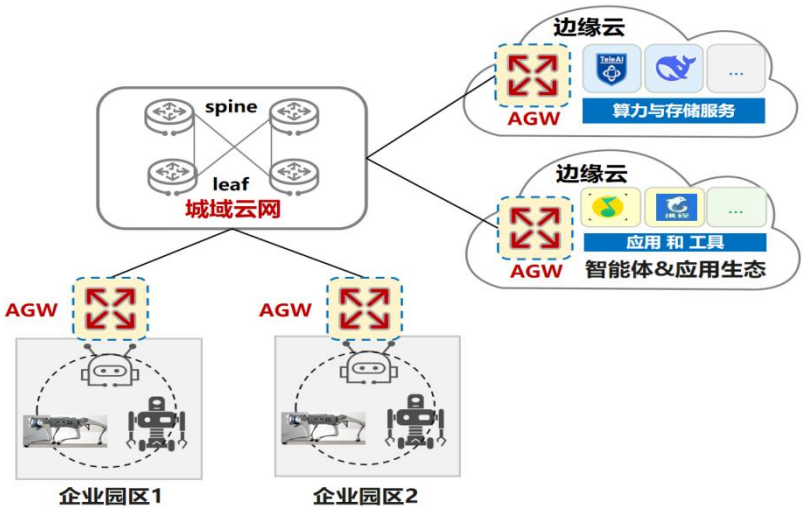


图 5 企业网关升级为智能体网关

企业网关长期部署于网络边界，是所有内外流量的必经节点和策略执行中枢，天然承担着路由转发、NAT 地址转换、VPN 隧道终结、SSL/TLS 卸载、应用层防火墙策略控制及精细化 QoS 调度等核心网络功能，这种“第一跳”位置与深度网络控制能力，使其天然成为承载 AGW 通信管理、身份寻址与策略协同功能的物理载体；同时，企业网关已在多年商用中构建起成熟稳定的身份认证与访问控制体系，普遍支持 RADIUS、LDAP/AD 对接、证书认证、SAML 单点登录及细粒度 ACL 策略引擎，并深度集成于企业 IT 治理体系之中，这为 AGW 所必需的 Agent ID 全生命周期管理、基于数字身份的强认证与动态授权、跨组织策略协同等安全能力提供了可复用的信任基础设施。

因此，基于企业网关升级实现 AGW 并非功能叠加式的简单改造，而是在已有网络控制平面之上，以模块化方式注入智能体身份管理、寻址服务与语义化策略框架，是一种充分利用既有投资、最大限度降低部署风险、确保策略一致性与运维连续性的平滑演进路径，让智能体通信能力扎实扎根于企业网络主干，成为

可管、可控、可审计、可持续演进的生产级基础设施。

- **典型场景分析：终端智能体跨园区协同**

在该场景下，基于企业网关升级的部署模式，依托企业网关已深度集成的 RADIUS/LDAP/AD 身份认证体系与细粒度 ACL 策略引擎，实现跨园区智能体的统一身份锚定与策略继承；通过将 SSL/TLS 卸载、VPN 隧道终结及应用层防火墙能力扩展至智能体通信层面，为企业各园区智能体构建双向可信通信通道，并支持基于设备证书、服务签名与策略约束的多轮意图协商；所有跨园区状态同步、能力通告与联合决策指令，均在企业网关策略框架内完成端到端加密与权限裁决，智能体间基于统一编码对齐语义空间，从而保证原始业务数据全程保留在本地系统，仅交换脱敏特征与结构化结果。

- **典型场景分析：企业智能体端云协同**

在该场景下，基于企业网关升级的部署模式，依托企业网关作为内外流量必经节点的“第一跳”位置，实现端侧智能体与云端智能体间的策略级协同；通过将企业网关的 NAT 地址转换、精细化 QoS 调度与 SSL/TLS 卸载能力，与云端智能体能力目录动态绑定，使端侧智能体在发起任务请求时，由企业网关自动完成语义解析、可信服务匹配与专属隧道建立；所有上传至云端的数据均经企业网关预处理为脱敏特征向量，云端返回的结果包则由企业网关校验其数字签名、字段策略与时效性后，再注入园区本地知识库或下发至执行终端。

4.4 基于云内资源池构建智能体网关

云内智能体网关应以 Region 为单位独立构建。云内每个 Region 在自有资源池中划出轻量计算节点（VM 或容器均可），部署一套独立运行的网关服务。租户智能体在启动时，通过标准 API 向本 Region 智能体网关发起一次注册请求，提供自身能力描述和访问端点，即完成接入。智能体网关随后为其分配唯一 Agent ID，并签发统一 CA 背书的 VC 凭证。所有后续智能体间调用，仍走原有 gRPC 或 HTTP 协议，网关在流量路径上自动完成 VC 校验、Token 计数、语义寻址匹配与基础策略拦截。租户完成注册后，其智能体能力会自动同步至本 Region 的智能体能力目录，并按标准字段结构化存储；任何其他智能体可向智能体网关发起一个标准查询请求，实时获得匹配的可用智能体列表与端点地址。

- **典型场景分析：云间异构智能体协同**

在该场景下，基于云网关升级的部署模式，依托各 Region 智能体网关独立运行、能力自治的特性，实现本 Region 内智能体能力的全量注册、标准化描述与实时目录同步；通过智能体能力目录与 DHT 等机制，将各 Region 能力元数据按租户策略、安全等级与业务域进行选择性共享，在保障数据主权前提下构建跨云智能体能力视图；所有智能体能力均以结构化字段发布，实现基于语义标签的高效匹配与多维策略过滤，确保调用方仅能发现并连接符合其业务约束与合规要求的可用目标。并且，所有凭证签发、调用日志、策略哈希均留存于各自 Region 本地，经标准化字段上报至集团 NaaS 平台，确保云间协同调用全程可验证、可追溯、可归属。

第五章：总结与展望

本白皮书聚焦智能体通信的核心使能需求，系统梳理了 ToH、ToB、ToC 三大典型场景下的差异化通信特征，深入剖析了身份可信、语义寻址、确定性 SLA、Token 级调度等关键挑战，并提出以“智能体网关为枢纽、Overlay/Underlay 双平面协同为底座、NaaS 能力开放为接口”的城域云网演进架构。

智能体互联网的规模化发展，离不开网络基础设施的深度适配与持续进化。我们希望通过本白皮书的研究与分析，激发更多业界同行的参与和讨论，期盼与众多合作伙伴一起携手并进，共同打造可信任、可寻址、可保障、可运营的智能体通信承载网络，为我国人工智能高质量发展构筑坚实网络基座。



媒体合作：contact@nida-alliance.org
工作机会：contact@nida-alliance.org
业务合作：contact@nida-alliance.org
官方网站：www.nida-alliance.org