Network Management Research Group
Internet-Draft
Intended status: Informational
Expires: 3 September 2026

C. Guo
China Mobile

March 2026

Agentic Network Architecture and Protocol for Supporting Agent Interconnection Communication and Multi-level Inference

draft-chuyi-nmrg-agentic-network-inference-00

Abstract

With the advent of the era of AI large models and intelligent agents, more and more scenarios about agent interconnection have emerged, such as collaboration among multiple agents within a household, intelligent robots cooperating to complete pipeline tasks in different operations of the industrial Internet, drone groups, intelligent vehicle networking, etc. These scenarios not only require low latency and high bandwidth, but also demand efficient information exchange and cross-domain coordination and scheduling capabilities in complex collaborative tasks. Therefore, new orchestration and management technologies and frameworks are needed in existing networks to address this. The interconnection of different agents also brings about an emergence of inference, with a large number of inference requests being processed from the mobile phone side to the cloud. In order to improve inference efficiency, in a cloud-edge-end multi-layer inference architecture, large models and agents at different levels cooperate to complete tasks, resulting in a complex intelligent agent interconnection network. Gateways and routers serve as forwarding entries on the network road highways, responsible for building communication channels for the agents spread throughout the network, which requiring function upgrades to support the continuously evolving inference service in the future.

Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction 3
2. Abbreviations and Definitions 3
3. Use Case 4
  3.1. Industrial Internet 4
  3.2. Smart Home 4
  3.3. Cloud-edge-end Collaboration Scenario of Intelligent Vehicle Networking 4
4. Definition and Functions of Agentic Gateway 4
  4.1. Introduction and Definition 4
  4.2. Intelligent Forwarding and Routing 5
  4.3. Intent and Network Environment Perception 6
  4.4. Protocol Compatibility and Conversion 6
  4.5. Support Multi-level Inference 6
  4.6. Information Management and Control 7
  4.7. Equipment-level Operation and Maintenance Autonomy . . . 7
  4.8. Safety 8
5. Architecture of Agentic Network for Multi-level Inference . . 8
  5.1. Architecture and Functions 8
6. IANA Considerations 8
7. Security Considerations 8

8. References 9
8.1. Informative References 9
8.2. Normative References 9
Author’s Address 9

1. Introduction

Agentic Network refers to a network composed of ubiquitous agents, intelligent network elements, and network resources. It addresses the issue of multi-agent collaboration across network domains, leveraging network capabilities to meet the interconnected communication and management needs of various forms of agents. It connects edge communication networks and devices, completing network intelligence upgrades. Toward the future, the Agentic (IP) Network will become a new generation of intelligent IP network, building a high-speed pathway for intelligent agent interconnection and serving as the core brain and transit station for agent interconnection and collaboration.

In this background, intelligent and autonomous network routing management &control will enter a new stage characterized by "intention-driven, autonomous collaboration, and inherent security". With intelligent agents as the core, the routing architecture will be restructured, achieving a qualitative change from "passive adaptation" to "active prediction" and from "single-point optimization" to "global collaboration".

2. Abbreviations and Definitions

Agentic (IP) network: A novel network architecture composed of ubiquitous ubiquitous agents, intelligent and non-intelligent network elements, computing resources, and network link resources, which supports agent interconnection technologies and functions such as intention transfer, semantic communication, knowledge and context-driven mechanisms, implementing agent interconnection protocols, and enables efficient collaboration among multiple agents as well as distributed multi-level inference.

Agentic gateway: Agentic gateway is the intelligent network element, capable of autonomously perceiving information and taking corresponding actions. It can perform all the functions of a gateway while integrating AI capabilities to execute planning, analysis, decision-making, and action execution for specific functions (such as intelligent route recommendation) or autonomous events.

(To be added)

3. Use Case

This chapter lists three typical application scenarios of agent interconnection, describing the capabilities and requirements of the agentic network and gateway needed.

3.1. Industrial Internet

Industrial Internet involves collaboration among smart manufacturing robots. In the production line, they cooperate to complete tasks, conduct multimodal data sensing, and fulfill their missions under the coordination and control of edge gateways. Edge gateways aggregate data from various device agents, perform local preprocessing (filtering invalid data, extracting key features), and avoid network congestion caused by directly connecting massive data to the cloud. At the same time, they can issue some control instructions (such as regional collaborative control instructions to achieve device linkage operation). Different agent devices within the gateway can share data, identify anomalies, and provide collaborative early warning and handling of faults.

3.2. Smart Home

(To be expanded)

3.3. Cloud-edge-end Collaboration Scenario of Intelligent Vehicle Networking

(To be expanded)

4. Definition and Functions of Agentic Gateway

4.1. Introduction and Definition

Intelligent network elements refer to network equipment hardware that integrates AI technology to fulfill network functions. Intelligent gateways leverage the analysis and generation capabilities of AI large or small models to accomplish new forwarding, control, management, and other functions in the context of the development in the AI era. These functions include but are not limited to intelligent traffic identification, intelligent route recommendation, intelligent forwarding, user-level/service-level service identification, autonomous operation and maintenance, achieving triple perception and tuning of themselves, the network, and services, while autonomously allocating resources, realizing event self-looping, supporting emerging inference services, and enhancing user service experience.

Each intelligent network element (NE) itself is an agent capable of performing all the functions of a NE. It has evolved from passive execution of single rules to active perception, discovery, and intelligent processing. Equipped with intelligent agent capabilities such as perception, planning, analysis, decision-making, and execution, it can do more than traditional NEs. As a "neuron" in the interconnected intelligent agent network, it can participate in higher-level decision-making and regulation.

Agentic gateway can support:

* The original gateway/router network elements were mostly passively responsive, executing a single rule function. The upgraded network elements can automatically complete functions in a closed loop, leveraging the capabilities of AI models, both large and small, to enhance automation and intelligence levels, reducing manual intervention.

* For agent interconnection, in order to face the complex agent interconnection scenarios in the future, it is necessary to enhance network capabilities, and the analysis, generation, and decision-making abilities of intelligent agents can be improved.

* The interconnection and management scheduling of multiple intelligent network elements. The advent of the agent era demands closer cooperation and interconnectivity among networks. The intelligence of network elements signifies the flexibility in supporting service operations.

4.2. Intelligent Forwarding and Routing

Unlike traditional address-based peer-to-peer information routing and forwarding, in the era of agent communication, tasks are first decomposed and grouped, followed by necessary communication for task execution or inference. After task triggering, agents and routing protocols in appropriate domains are matched based on service characteristics and requirements, enabling intelligent agent routing decisions and addressing.

Service characteristics can be based on the current task classification (such as ordering, navigation, etc.), or can be classified according to time-delay sensitivity, data without loss, high bandwidth, etc.

The capability graph can be used to search for an appropriate list of agents during addressing. This process can be combined with the capabilities of agents, application intents, real-time loads, and link quality for dynamic addressing. Based on the characteristics of the agents, targeted traffic distribution can be implemented. Furthermore, through self-learning, the selection of the optimal path can be reinforced.

4.3. Intent and Network Environment Perception

Agentic gateway should achieve autonomous event detection and handling, perceive and forward information intentions, and sense the state of the intelligent agent (online/offline), load, link quality, computing power, and other environmental states. Based on this, it should search for the next hop/destination. At the same time, it must also perceive abnormal signals, understand the significance of collected data signals, trigger corresponding actions, possess key technologies such as strategy translation and generation, achieve triple perception and tuning of itself, the network, and the business, autonomously allocate resources, and realize an event self-loop.

4.4. Protocol Compatibility and Conversion

* Regardless of whether the devices come from the same manufacturer, a unified protocol must be adopted to facilitate routing to nearby or edge large model servers within the same gateway. The gateway can automatically identify the protocol, align fields, semantics and capability descriptions, then generate adaptations.

* Supports semantic communication, supports the conversion of modalities between different agents, and supports the conversion of context content (such as MCP result conversion).

4.5. Support Multi-level Inference

In a cloud-edge-end multi-tier inference architecture, the gateway connects end-side agents, edge nodes, backbone routers, and the cloud. The edge resource pool typically deploys lightweight or specialized large models and agents, while the central cloud can host super agents and full-scale large models. Before cross-domain interconnection, the gateway performs edge analysis and processing to determine which tasks should be sent to edge nodes for inference and which should be forwarded to central nodes. This enables hierarchical forwarding of service flows and data pre-processing, avoiding the upload of massive amounts of raw redundant data to the cloud. This analysis process can leverage large and small model capabilities to complete the tasks.

4.6. Information Management and Control

Discovering newly online agents and reporting this information to higher-level management, allowing registration and reporting through the edge gateway. It manages intra-domain agent addresses, synchronizes and maintains teaming information, and ensures consistency between routing identifiers and networking relationships.

Acting as an execution entity for upper-layer orchestration and management, it possesses information reporting capabilities, executes management instructions from higher layers, and implements cross-domain access control. The information reported by network elements is aggregated at the management layer, where it is uniformly invoked and analyzed by the super agents in the orchestration layer.

4.7. Equipment-level Operation and Maintenance Autonomy

Equipment-level operation and maintenance autonomy can achieve autonomous and fully automated Operation and Maintenance through agentic network elements:

* Capabilities of Autonomous Perception and Problem Handling: Equipped with the ability to autonomously perceive the environment and detect data issues, identify problems, and handle network events. It can independently complete fault diagnosis and problem localization, support automatic duty reporting, and enable automatic configuration distribution along with self-inspection before service deployment.

* Autonomous Event Handling: Capable of real-time perception of the network environment and business data, automatically completing data reporting to achieve 24/7 unattended autonomous operation. It proactively discovers, identifies, and processes network events, accurately determines the type of issue, and autonomously completes the analysis and resolution of network events

* Interoperation with Other Intelligent Elements: To meet specific operation and maintenance requirements for users, it can autonomously initiate mutual recognition, interconnection, mutual inspection, and information exchange with other intelligent network elements. It supports interconnection and interoperation with other agents, mutual status verification, and efficient exchange of fault information.

4.8. Safety

* Conducting network security status monitoring, with intelligent threat tracing and full lifecycle auditing capabilities, leveraging AI-driven situational awareness and anomaly detection to identify covert attacks.

* Access control (blacklists, etc.)

* Data security

(To be expanded)

5. Architecture of Agentic Network for Multi-level Inference

5.1. Architecture and Functions

Agentic (IP) network refers to a network composed of software and hardware designed for ubiquitous agent interconnection. It primarily consists of underlying connectivity resources, computing resources (computing nodes, data centers, etc.), intelligent agent-enabled network elements in the middle, an upper-layer coordination management system, as well as the applications at the edge and in the center. It is deployed in a master-slave distributed mode and supports cross-domain collaboration.

The architecture can include the application layer, orchestration layer, management and control layer, device and network layer, computing resource layer, and data layer. The application layer primarily targets agent applications and scenarios. The capabilities of the management & control layer and orchestration layer can be realized through a unified platform. The device and network layer focuses on existing network devices. The computing resource layer refers to the distribution of computing resources that facilitate agent interconnection and inference. The data layer provides a unified shared storage for data.

(To be expanded)

6. IANA Considerations

This document has no requests to IANA.

7. Security Considerations

This document describes architecture and protocol of agentic network. As such, the following security considerations remain high level, i.e., in the form of principles, guidelines or requirements.

8. References

8.1. Informative References

[CCSA-AIWan]
CCSA, "Research Report on Artificial Intelligence Wide Area Network (AI WAN) (2025)", June 2025.

8.2. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://www.rfc-editor.org/info/rfc2119>.

Author’s Address

Chuyi Guo
China Mobile
Beijing
100053
China
Email: guochuyi@chinamobile.com