Network Working Group
Internet-Draft
Intended status: Informative
Expires: 8 January 2026

M. Han, Ed.
N. Zhang, Ed.
X. Gu, Ed.
China Unicom

7 July 2025

The Impact of AI Agent to Network Infrastructure
draft-han-ai-agent-impact-infra-00

Abstract

This document disucss and analyses the impact of AI Agent on network infrastructure aiming to facilitate the discussion and standardization about AI Agent communication within IETF.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at https://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on 8 January 2026.

Table of Contents

1. Introduction

The rapid advancement of AI agents has introduced transformative changes across various domains, particularly in network infrastructure. AI agents, the autonomous, intelligent entities capable of perception, reasoning, and decision-making, are increasingly being deployed in diverse scenarios. On the one hand, the network needs to provide communication support for AI agent collaboration, such as industrial, office, smart home and other scenarios. On the other hand, AI agents can be deployed in the network to achieve automatic network O&M.

Traditional network infrastructures were primarily designed for human-driven communication (e.g., web browsing, video streaming). However, the rise of multi-agent systems, where AI agents collaborate or compete in real-time, demands ultra-low-delay communication, dynamic resource allocation, and enhanced security protocols.

Besides, AI agents are revolutionizing network O&M. AI agents, powered by reinforcement learning and large language models, enable autonomous fault detection, predictive maintenance, and intent-driven networking.

This draft explores the impact of AI Agents on network infrastructure, with a focus on Network for Agent Communication and Agent-based Network O&M.

2. Conventions and Definitions

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here. Abbreviations and definitions used in this document:

3. Network for Agent Communication

3.1. Scenarios

3.1.1. Agent Communication in LAN

a. Office networks In the office network scenario, agents rely on a stable LAN environment to build an efficient collaboration system. Typical office agents, such as personal AI assistant, can automatically analyze users’ work requirements, such as writing emails, organizing reports, and quickly generating drafts. In the situation of team collaboration, multiple agents share data in real-time, automatically arrange task priorities, which can reduce manual communication costs and significantly improve work efficiency. There is a requirement for office network to provide security guarantees for data interaction among agents to prevent the leakage of sensitive information.

b. Industrial networks In the industrial scenario, the LAN establishes a low-delay and highly reliable communication environment for agents, enabling efficient interconnection among production equipment, logistics robots, and management systems. For example, in the warehousing and logistics process, the agents carried by AGV communicate with the central dispatching agent through the network, dynamically planning transportation routes based on real-time traffic conditions to improve the efficiency of material distribution.

3.1.2. Agent Communication across WAN

a. Agents directly access to public networks When agents directly access public network, they can achieve efficient coordination over a wide range and across regions. For example, in smart city, traffic management agents, environment monitoring agents, and public service agents collaborate efficiently through public networks. However, the openness of public networks also brings challenges in data security and privacy protection, which require the use of technologies such as encryption and authentication to ensure the security of agent communication.

b. Agents in Campus/DC to communicate across the WAN AI agents within different DCs can achieve resource sharing and remote collaboration through WAN. For example, when a data center experiences a shortage of computing resources, the resource scheduling agent in this DC negotiates with resource scheduling agents of other DCs through the WAN to divert some tasks to DCs with idle resources. This requires the WAN to ensure low delay.

3.1.3. Agent Communication in Fields

a. Interaction through 3GPP radio networks In the office network scenario, agents rely on a stable LAN environment to build an efficient collaboration system. Typical office agents, such as personal AI assistant, can automatically analyze users' work requirements, such as writing emails, organizing reports, and quickly generating drafts. In the situation of team collaboration, multiple agents share data in real-time, automatically arrange task priorities, which can reduce manual communication costs and significantly improve work efficiency. There is a requirement for office network to provide security guarantees for data interaction among agents to prevent the leakage of sensitive information.

b. Interaction through WiFi In the industrial scenario, the LAN establishes a low-delay and highly reliable communication environment for agents, enabling efficient interconnection among production equipment, logistics robots, and management systems. For example, in the warehousing and logistics process, the agents carried by AGV communicate with the central dispatching agent through the network, dynamically planning transportation routes based on real-time traffic conditions to improve the efficiency of material distribution.

c. Interaction through satellite networks In the industrial scenario, the LAN establishes a low-delay and highly reliable communication environment for agents, enabling efficient interconnection among production equipment, logistics robots, and management systems. For example, in the warehousing and logistics

process, the agents carried by AGV communicate with the central dispatching agent through the network, dynamically planning transportation routes based on real-time traffic conditions to improve the efficiency of material distribution.

3.2. Gap Analysis

3.2.1. Agent ID management among heterougenous networks

Agent ID management in heterogeneous networks refers to the processes and systems for uniquely identifying, authenticating, and managing agents (software entities, devices, or services) across networks with different architectures, protocols, and technologies. Key challenges include interoperability, scalability, security, dynamicity, and privacy protection.

3.2.2. Mobility Management (esp. for Non-3GPP networks)

The AI agent era exacerbates existing gaps in mobility management, particularly in interoperability, security, and context awareness. Addressing these requires,

AI-native mobility frameworks that prioritize agent intent over device-centric logic. Dynamic, secure abstraction layers for cross-protocol interoperability. Standardized interfaces for agents to share mobility intent with networks. Energy-efficient, role-based QoS models tailored to autonomous AI agents.

3.2.3. Pervassive Point-to-Point Secure Channel

Pervasive P2P secure channels in the AI agent era require a paradigm shift from device-centric, static security to agent-aware, dynamic, and scalable frameworks. Addressing these requires,

Developing lightweight, context-aware trust models for ephemeral agent interactions. Implementing distributed key management and low-latency protocols for massive agent swarms. Creating adaptive security abstractions that work across heterogeneous networks. Integrating privacy-preserving techniques, such as metadata obfuscation and fine-grained access control, into core channel design. Establishing AI-specific standards that align security with agent roles and autonomy.

3.3. Potential new works

3.3.1. Pervasive Mobility Managment in fixed networks

TBD

3.3.2. On-demand Point-to-Point Private Line service

TBD

4. Agent-based Network O&M

4.1. Scenarios

4.1.1. Autonomic O&M through AI Agent built into Devices

In traditional congestion control method, there is a lack of cross-device historical data sharing and AI model collaboration between devices. It is unable to adaptively optimize based on real-time traffic patterns. When AI agents are introduced into network devices, intelligent collaboration can be achieved. Devices synchronize real-time link bandwidth and TOP-N traffic characteristics through BGP extension. The AI agent dynamically define congestion thresholds based on traffic data, replacing manual threshold configuration. Upon detecting congestion, devices negotiate AI-generated policies (e.g., dynamic adjustment of Multi-Exit Discriminator (MED) values) and route traffic precisely to lightly loaded links. Reinforcement learning is applied to dynamically optimize policy parameters during this process.

4.1.2. Agent interactions between Devices and Controllers

In the current IPv6 end-to-end traffic monitoring scenario, traffic data collection and analysis rely on manual intervention, while the large volume of live network traffic data results in high resource requirements. When AI agents are deployed in network controllers and devices, intelligent collaboration can be achieved between controllers and edge devices. The controller agent collects and monitors IPv6/IP traffic data in real time, and performs preliminary analysis including flow pattern recognition and IPv6/IPv4 traffic ratio trending. The device agent collects customer traffic data, decomposes traffic distribution characteristics to identify high-value business scenarios, and synchronizes these insights to the controller agwnt. The controller agent integrates global traffic ingress/egress data to construct regional traffic matrices, obtaining analysis results such as traffic distribution and link utilization.

4.2. Potential new works

4.2.1. Extension of Anima Framework & Protocols

Considering the Anima Framework & Protocols [RFC7575] with AI Agent, which provides more intelligent operation and management of network devices to achieve the Intention-driven network and Auto-driven network.

4.2.2. South-bound Interface Extension for Agentic O&M

The relatted agent communication protocols such as MCP, A2A etc., need to be extended to support the collabration between network devices and controller.

5. Security Considerations

TBD

6. IANA Considerations

TBD

7. References

7.1. 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>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

7.2. Informative References

[RFC7575] Behringer, M., Pritikin, M., Bjarnason, S., Clemm, A., Carpenter, B., Jiang, S., and L. Ciavaglia, "Autonomic Networking: Definitions and Design Goals", RFC 7575, DOI 10.17487/RFC7575, June 2015, <https://www.rfc-editor.org/info/rfc7575>.

Authors' Addresses

Mengyao Han (editor)
China Unicom
Beijing
China
Email: hanmy12@chinaunicom.cn

Naihan Zhang (editor)
China Unicom
Beijing
China
Email: zhangnh12@chinaunicom.cn

Xinrui Gu (editor)
China Unicom
Beijing
China
Email: guxr12@chinaunicom.cn