

Network Working Group
Internet-Draft
Intended status: Informational
Expires: 16 September 2026

Z. Du
China Mobile
15 March 2026

Routing Considerations in Agentic Network
draft-du-catalist-routing-considerations-01

Abstract

As the development of the AI technology, an AI Agent would be able to do some tasks as an assistant to human beings. During the task process, the Agent may need to connect to other Agents with different skills relative to the task. The Agent to Agent communication is a new kind of traffic for Internet, and some new requirements for networking are proposed. This document describes some routing considerations in the agentic network, especially for the cross-domain scenarios, in which the agentic network works as an overlay network above the IP network.

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

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 16 September 2026.

Copyright Notice

Copyright (c) 2026 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Revised BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Revised BSD License.

Table of Contents

1. Introduction	2
2. Intra-domain and Inter-domain	3
3. Peer-to-peer Connection and Message System	3
3.1. Direct-Connected Mode	4
3.2. One Agent Communication Server Mode	4
3.3. AGW Network	5
4. Routing Considerations for AGW Network	5
4.1. Forwarding Based on Agent ID	6
4.2. Forwarding Based on AGW ID	6
4.3. Forwarding Based on Channel ID	6
4.4. Forwarding Based on Skill	7
5. IANA Considerations	7
6. Security Considerations	7
7. Acknowledgements	7
8. References	7
8.1. Normative References	7
8.2. Informative References	7
Author's Address	8

1. Introduction

In [I-D.rosenberg-ai-protocols], some use cases and requirements for AI Agent protocols are introduced. Meanwhile, a framework is described for the agent communications, and it includes the communications between AI agent and User, AI agent and API, AI agent and AI agent. In this document, we mainly focus on the Agent to Agent communication scenarios.

In the agentic network, it is assumed that many Agents exist, and they need to cooperate to complete tasks. The AI Agents in the same task group need to form a virtual network to communicate with each other, and after the task completing, the communication group may be released.

The main purpose of this document is to describe some routing considerations in the agentic network. Other issues, such as the discovery and authentication of the target Agents, are out of scope.

2. Intra-domain and Inter-domain

The Agent commutation nowadays mainly happens in the intra-domain scenarios, in which many problems would be simpler. As introduced in the [I-D.zyyhl-agent-networks-framework], the scenarios about the single trusted domain are analyzed firstly, and cross-domains are suggested to be considered in future. In that trusted domain, a Registration Server is responsible for establishing the trust relationship, and a Communication Server is responsible for communications between the Agents. For these intra-domain cases, most of the agents should be well known to the administrator, or to the Registration Server.

When the scenarios are extended into the Internet level, where cross-domains become unavoidable, things become more complicated. As mentioned in the [I-D.rosenberg-ai-protocols], an Agent should not trust another Agent only because it is one of the results provided by a search engine on the Internet, and a high level of trust is required for one AI agent to talk to another one in the inter-domain cases. Meanwhile, in the cross-domain scenarios, if different domains could follow the same standardized realization, the communication would be much easier.

3. Peer-to-peer Connection and Message System

There are many different tasks and different scenarios for Agent communications. In this section, we introduce three kinds of connection modes as follows.

1. Direct-connected mode: AI agents directly send and receive messages without the need for intermediate nodes for processing.
2. Indirect Communication by using one Agent Communication Server: It often appears in a single domain scenario. Communication between AI agents requires processing/relaying by an Agent Communication Server, and the AI agent must be aware of and interact with the Agent Communication Server.
3. Indirect Communication by using the AGW network: It often happens in cross-domain scenarios. An AI agent firstly connects a Communication Server or called an Agent Gateway (AGW), and then communicates with other Agents by using the AGW network. The Agents in the task group may attach to different Agent Gateways, so that more hops are involved.

The second case can also work for cross-domain scenarios, if we assume that a super Agent Communication Server exist. However, the distributed solution mentioned as the third one will have benefits such as the path length and the scalability.

3.1. Direct-Connected Mode

If the Agent communication only happens between two Agents, the peer-to-peer mode would be the most straightforward approach. We can also call it a direct-connected mode.

If more Agents are involved in the communication for the task, this direct-connected mode can also work. However, the leader of the task group needs to be responsible for all the transmitting of the messages.

Normally, the leader is the Agent that receives the task, and finds partners to complete the task. Thus, in this case, the leader initiates the task group, and is responsible for the interaction of the Agents, while other Agents only need to have a connection to the leader.

3.2. One Agent Communication Server Mode

As mentioned in [I-D.mpsb-agntcy-messaging], the Advanced Message Queuing Protocol (AMQP) is often used for enterprise messaging systems. With the message system, asynchronous processing is enabled, and reliable message delivery for distributed transactions can be guaranteed. Additionally, it supports flexible routing and easy integration of new components, enhancing overall architecture resilience and adaptability.

On the other hand, as mentioned in [I-D.nandakumar-ai-agent-moq-transport], the Media over QUIC Transport (MoQT) provides an alternative for real-time agent communication systems. The MoQT relay infrastructure provides native publish/subscribe messaging capabilities with efficient real-time data streaming and built-in prioritization. Crucially, its integrated caching mechanism ensures robustness by enabling late-joining agents to receive recent context/messages, facilitates recovery from transient disconnections, and allows efficient replay of operation history. Furthermore, it supports hierarchical namespace organization and protocol multiplexing, which significantly enhances scalability and interoperability across diverse agent ecosystems. In both approaches, the communication architecture is simplified by centralizing complex functions within the Agent Communication Server. For AMQP-based systems, agents connect to a

message queue server that handles routing and delivery logic. Similarly, in MoQT-based systems, agents establish QUIC connections to the MoQT Relay, which manages routing, discovery, and quality-of-service mechanisms.

In these cases, an Agent only needs to have a connection to the Agent Communication Server. We can realize the message distribution system in this Agent Communication Server.

3.3. AGW Network

For cross-domain scenarios, multiple Agent Communication Servers could exist. As it is for inter-domain, we can also consider that Agent Communication Server works as the Agent Gateways.

It is assumed that each domain has an Agent Gateway, and the Agent Gateways have already been connected by some means. Thus, the A2A connection is composed of three parts, i.e., from the source Agent to the AGW A1 that it attaches to, from AGW A1 to another AGW A2, and from AGW A2 to the target Agent.

The structure is described in [I-D.mpsb-agntcy-slim], in which the Agent Gateway is considered as a Message Node or a Routing Node.

4. Routing Considerations for AGW Network

After connecting the Agent Gateways, we establish an overlay network above the IP network for cross-domain scenarios of the Agent communication. Based on this overlay network, Agents can dynamically establish virtual groups on demand.

The AGWs should be able to support flexible forwarding mechanisms. Four requirements are listed as follows.

1. Forwarding based on Agent ID: AI agents should have a structured ID that can be discovered and addressed.
2. Forwarding based on AGW ID: The forwarding table of AGW IDs should be pre-configured in the AGW.
3. Forwarding based on Channel ID: The channel ID is related to the virtual group. It supports multicast in the task group.
4. Forwarding based on Skill: It happens in the discovery stage. When many agents can provide the same skill, they can advertise the same skill ID into the AGW network.

For the last case, the registration/discovery system can also give the suggestion about which Agent should be connected. However, anycast on the AGW network can also be a potential solution for some scenarios. It is similar to the relationship between the GSLB (Global Server Load Balance) technology based on DNS and the anycast technology based on BGP.

4.1. Forwarding Based on Agent ID

The number of Agents in the agentic network could be very large. However, the Agent IDs normally can not be aggregated as IP addresses do. Thus, we do not need the AGW to have a complete forwarding table for all the Agent ID in advance.

If a target Agent ID is received, and an AGW does not know how to route the traffic. It can query the registration/discovery system for that ID.

Meanwhile, for the intra-domain traffic, the AGW could support the forwarding of the traffic more easily. The AGW should be aware of the Agents that attach to it, and maintain an attached Agent table.

4.2. Forwarding Based on AGW ID

As talked before, the interconnection of the Agent Gateway is the pre-condition of the AGW network. Some information exchanges should be supported among the AGWs.

If an AGW can not recognize a specific Agent ID, it should perform a discovery procedure, and obtain the target AGW ID that the Agent attaches to. Next, the AGW forwards the message to the target AGW, and the target AGW should be aware of the location of the Agent.

Specially, the Agent ID can be structured to include or imply the home AGW ID (e.g., agent123@agw-domain-a.example), then the lookup can be resolved to the target AGW efficiently.

4.3. Forwarding Based on Channel ID

As the communication happens among a task group, and the AGW network would support many virtual networks, each for a task group, AGW should support forwarding based on the channel ID.

AGWs that are involved in the traffic of a task group should maintain a forwarding table of the channel ID corresponding to the task group. The forwarding table can be built more efficiently if the Channel ID itself is structured to include the AGW ID.

Meanwhile, the AGW should be aware of the local-connected Agents that have subscribed the channel.

4.4. Forwarding Based on Skill

In the agentic network, many Agents own the same skill, and they can advertise the same target address into the agentic network. The selection method may related to the network distance, the predicted service experience, etc.

5. IANA Considerations

TBD.

6. Security Considerations

TBD.

7. Acknowledgements

TBD.

8. References

8.1. 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>>.

8.2. Informative References

[I-D.mpsb-agntcy-messaging]
Muscariello, L., Papalini, M., Sardara, M., and S. Betts, "An Overview of Messaging Systems and Their Applicability to Agentic AI", Work in Progress, Internet-Draft, draft-mps-b-agntcy-messaging-01, 24 February 2026, <<https://datatracker.ietf.org/doc/html/draft-mps-b-agntcy-messaging-01>>.

[I-D.mpsb-agntcy-slim]
Muscariello, L., Papalini, M., Sardara, M., and S. Betts, "Secure Low-Latency Interactive Messaging (SLIM)", Work in Progress, Internet-Draft, draft-mps-b-agntcy-slim-01, 24 February 2026, <<https://datatracker.ietf.org/doc/html/draft-mps-b-agntcy-slim-01>>.

[I-D.nandakumar-ai-agent-moq-transport]

Nandakumar, S. and C. F. Jennings, "MOQ Transport for Agent Protocols", Work in Progress, Internet-Draft, draft-nandakumar-ai-agent-moq-transport-00, 2 March 2026, <<https://datatracker.ietf.org/doc/html/draft-nandakumar-ai-agent-moq-transport-00>>.

[I-D.rosenberg-ai-protocols]

Rosenberg, J. and C. F. Jennings, "Framework, Use Cases and Requirements for AI Agent Protocols", Work in Progress, Internet-Draft, draft-rosenberg-ai-protocols-00, 5 May 2025, <<https://datatracker.ietf.org/doc/html/draft-rosenberg-ai-protocols-00>>.

[I-D.zyyhl-agent-networks-framework]

Zhouye, Yao, K., Yu, M., Han, M., and C. Li, "Framework for AI Agent Networks", Work in Progress, Internet-Draft, draft-zyyhl-agent-networks-framework-01, 20 October 2025, <<https://datatracker.ietf.org/doc/html/draft-zyyhl-agent-networks-framework-01>>.

Author's Address

Zongpeng Du
China Mobile
No.32 XuanWuMen West Street
Beijing
100053
China
Email: duzongpeng@foxmail.com