CAICT 中国信通院

AI based Network Management Agent (NMA): Concepts & Architecture

IETF 123

draft-zhao-nmop-network-management-agent/

zhaoxing@caict.ac.cn

Authors:

Xing Zhao (CAICT)

Minxue Wang (China Mobile)

Bo Wu (Huawei)

Daniele Ceccarelli (Cisco)

Jin Zhou (ZTE)

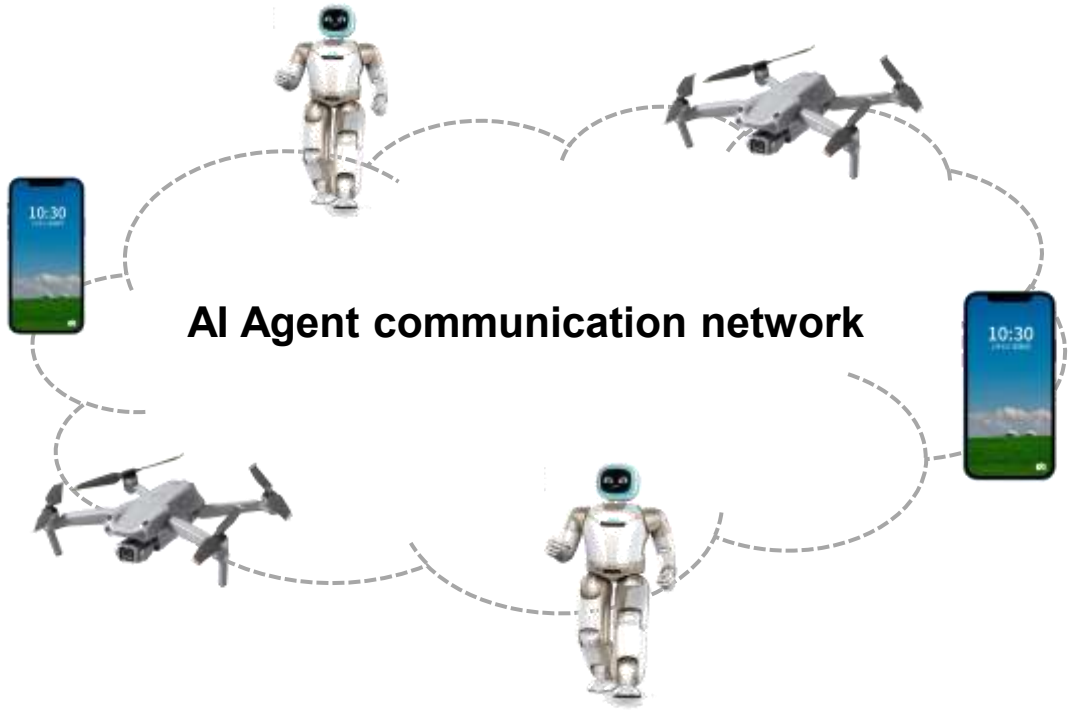
Haomian Zheng (Huawei)

Relationship between AI Agent and Network

- AI Agent has become one of the mainstream application forms for the implementation of large AI models.

Network for AI Agent

- **AI Agent communication network**
  - Including protocols, access authentication, etc.



AI Agent for Network

Scope of this draft

- **AI Agents** can be widely used in the operations to achieve an **E2E autonomous closed loop**, improving the intelligence of network management and control.

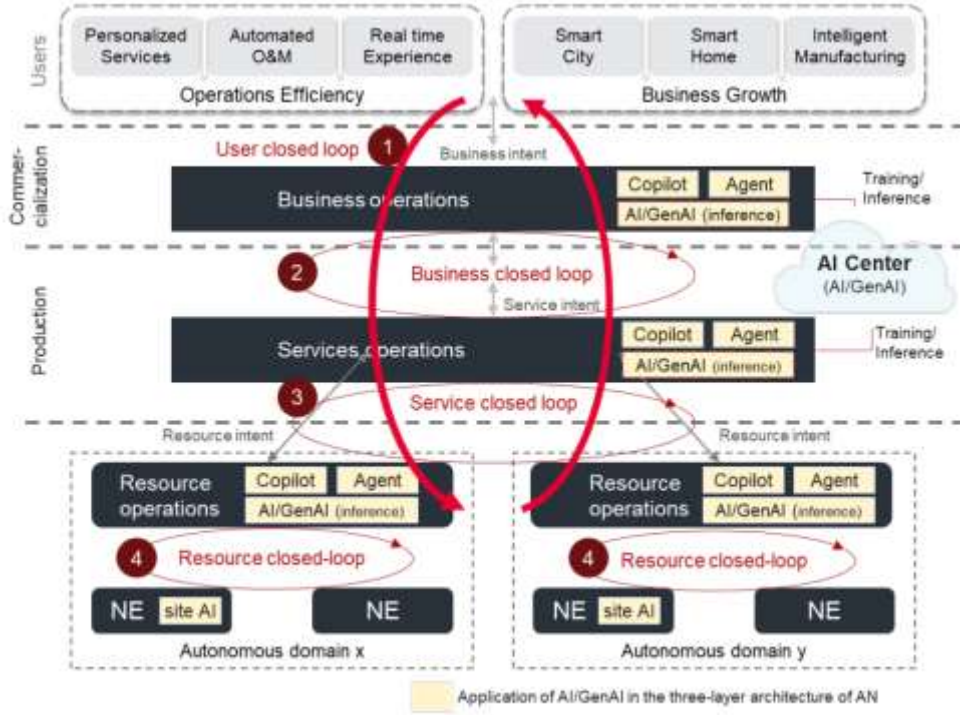


Figure 9: Full-stack AI / GenAI in the three-layer architecture of AN

TMF whitepaper "Level 4 industry blueprint - high-value scenarios" proposed the full-stack AI/GenAI in the three-layer architecture of AN

IETF standardization work on AI Agents for network

- **AI based Network Management Agent(NMA): Concepts and Architecture** (draft-zhao-nmop-network-management-agent)
  - Proposed in October 2024.
  - This document presents the concept of AI based network management agent(NMA), provides the basic definition and reference architecture of NMA, discusses the relationship of NMA with traditional network controller or other network management entity by exploring the deployment mode of NMA, and proposes the common processing flow and typical application scenarios of NMA.
- **Large Model based Agents for Network Operation and Maintenance** (draft-chuyi-nmrg-ai-agent-network-00)
  - Proposed in March 2025.
  - This document identifies typical scenarios requiring enhanced intelligence, and explains how AI Agents and large model technologies can empower networks to address operational pain points, reduce manual efforts, and explore impacts on network data, system architectures, and interfaces correspondingly.
- **A Framework for LLM-Assisted Network Management with Human-in-the-Loop** (draft-cui-nmrg-llm-nm-00)
  - Proposed in March 2025.
  - This document defines an interoperable framework that facilitates collaborative network management between Large Language Models (LLMs) and human operators.

Standardize the general framework and capabilities of the Agent, rather than defining how the agent is implemented.

The goal and scope of this draft

CAICT 中国信通院

- This draft is trying to **give a standardized common architecture of AI Agent for network management--->Network Management Agent (NMA).**
  - Provide the framework of **network management agent** (AI agent for network management) Clarify the relationship of NMA with existing controllers
  - Define **functional requirements** of NMAs for different scenarios and related interfaces

This work is focus on the common framework of control plane.

The specific agent implementation details are not in the scope of this draft.

Hope this draft can be the start of standardization work on AI Agents for network within the IETF

Updates since 00-version

1) Supplement the background content in **Section 1.1**

2) **Add Section 1.2** to optimize the conceptual definition of NMA

3) **Revise Chapter 3** significantly, including:

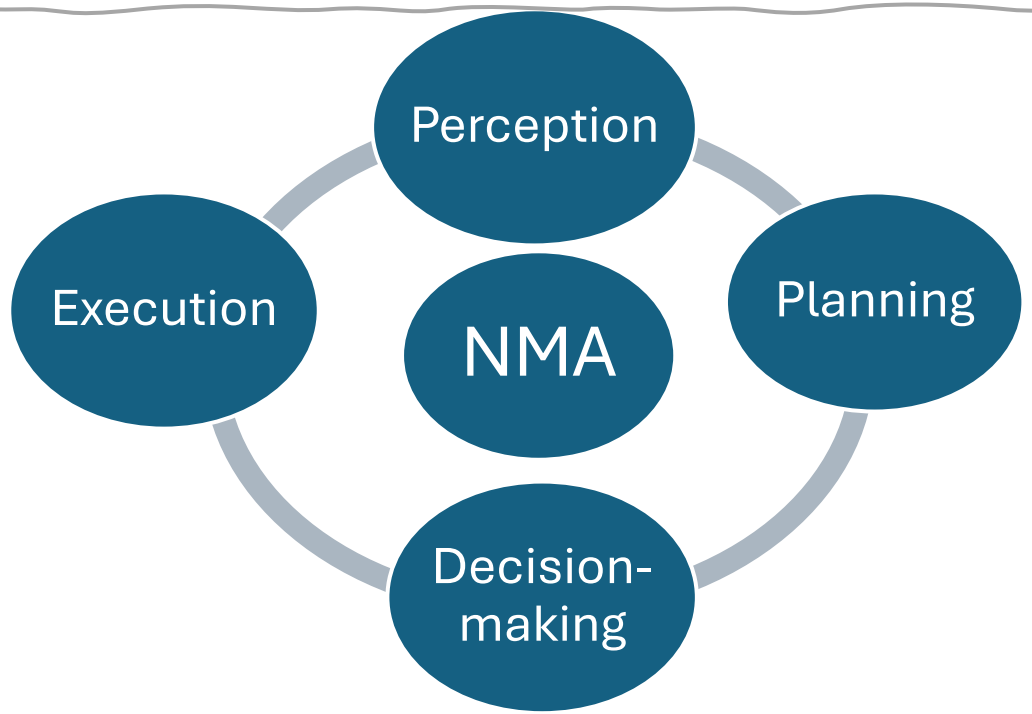
- ✓ Add Section 3.1 to describe the NMA-based intelligent network management and control framework, and explain the relationship of this draft with RFC 8969.
- ✓ Add Section 3.2 to describe the functional requirements of NMA.
- ✓ Revise Section 3.3 to streamline and optimize the definition of NMA's functional architecture.
- ✓ Add Section 3.4 to describe NMA's external interfaces.

4) Add Chapter 4 to elaborate on application considerations under single-Agent and multi-Agent scenarios.

Major updates since 00-version (1)

Optimize the definition of Network management agent (NMA) -→AI Agent for network management

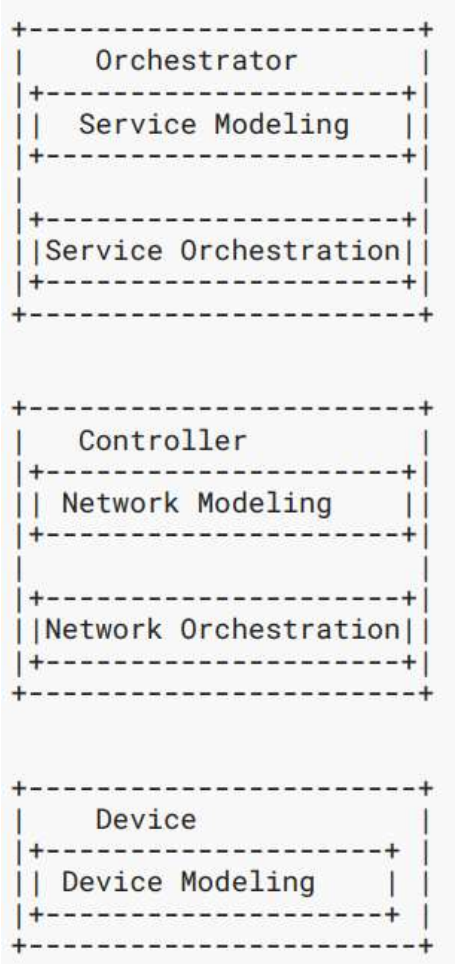
- ***A network management entity built based on ML/AI and equipped with the autonomous closed-loop task processing capabilities.***
- It can automatically carry out network status perception, task intent interpretation, task planning, decision-making and task execution operations based on user task intentions or preset goals, so as to achieve closed-loop processing of scenarios-oriented network management tasks.



Major updates since 00-version (2)

Add the Intelligent Network Management and Control Framework Based on NMA

- **Relation with RFC8969**: this draft is an enhancement of intelligent capabilities of RFC8969.
- **Position of NMA:** NMA can exist at both the Controller and Orchestrator levels. Some end-side AI components may be added on the device instead of a complete NMA.



RFC 8969: Layering and Representation within a Network Operator

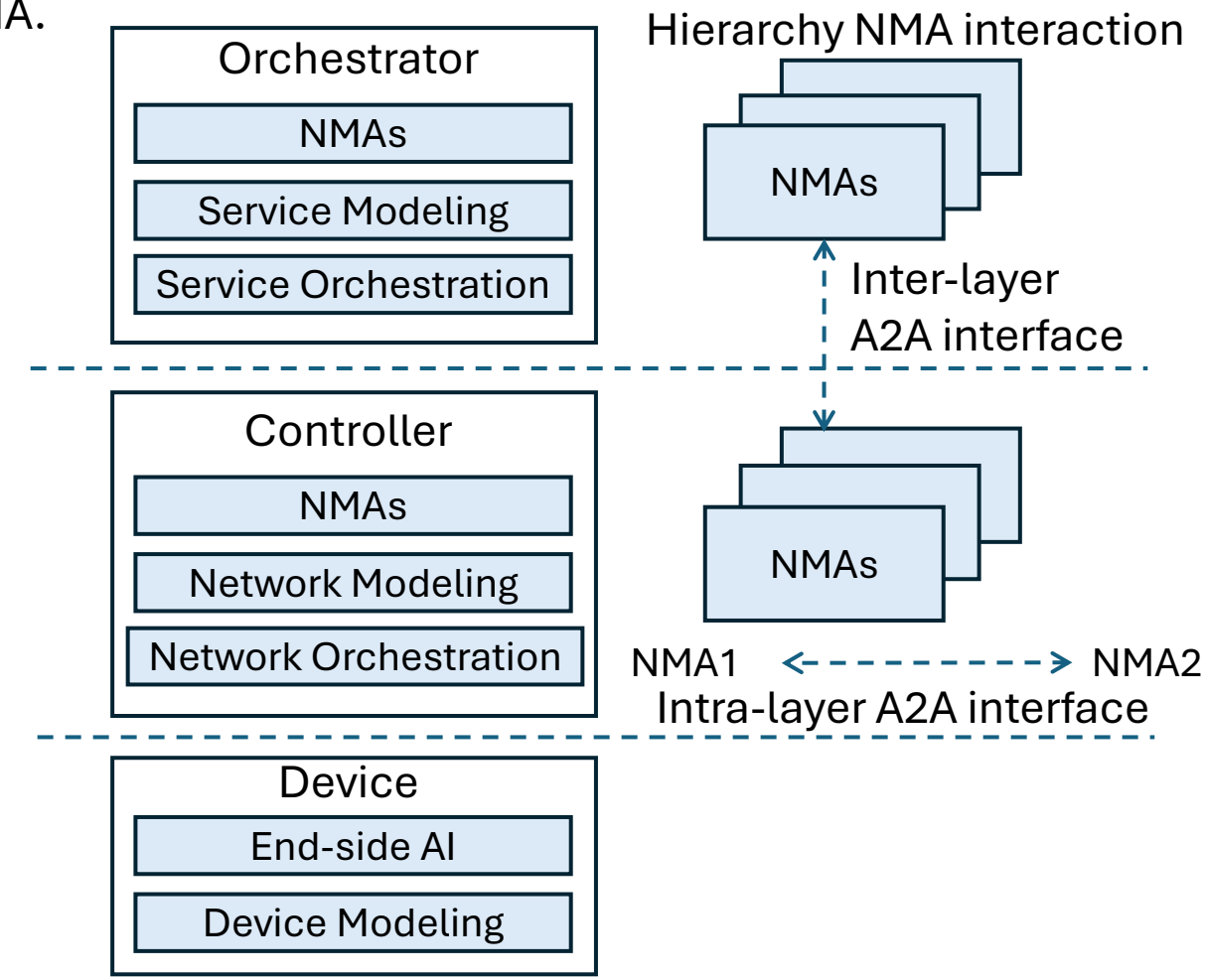
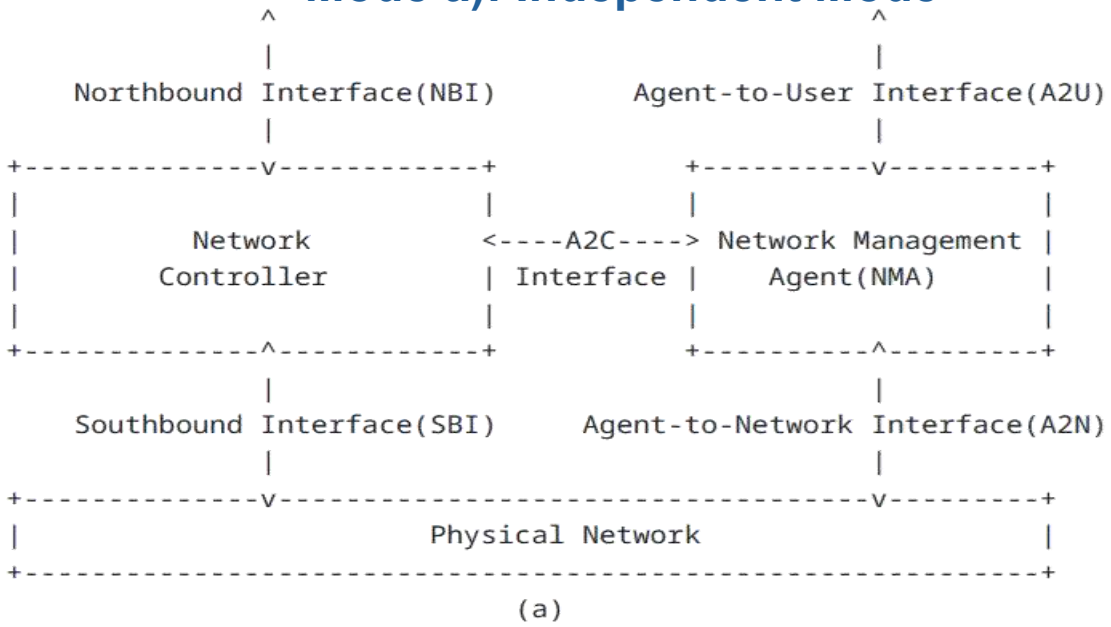


Fig.1 Enhanced intelligent network management and control framework based on NMA

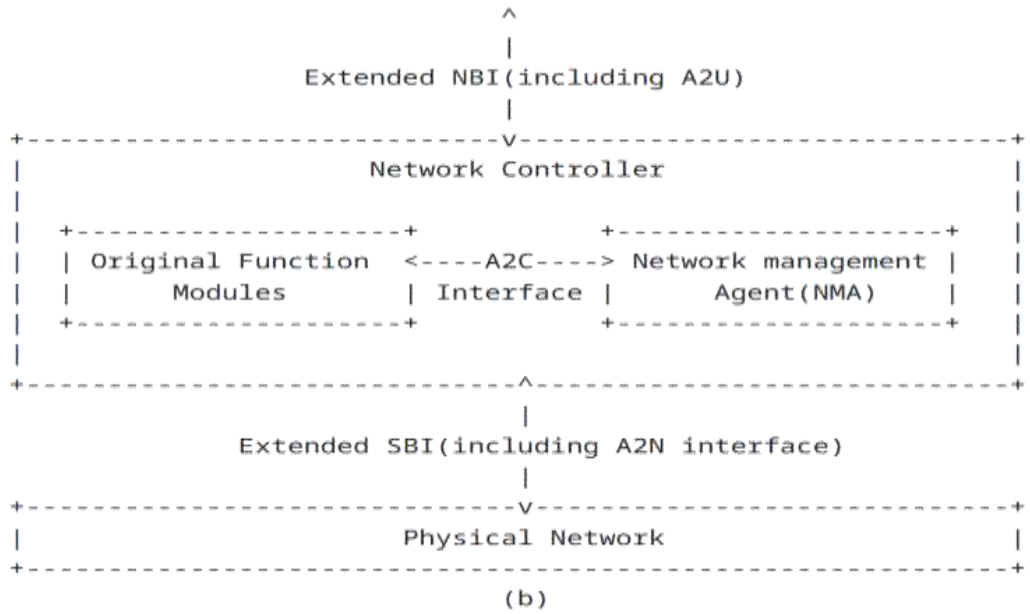
Major updates since 00-version (3)

Deployment modes of NMA:

Mode a): Independent Mode



Mode b): Integrated Mode

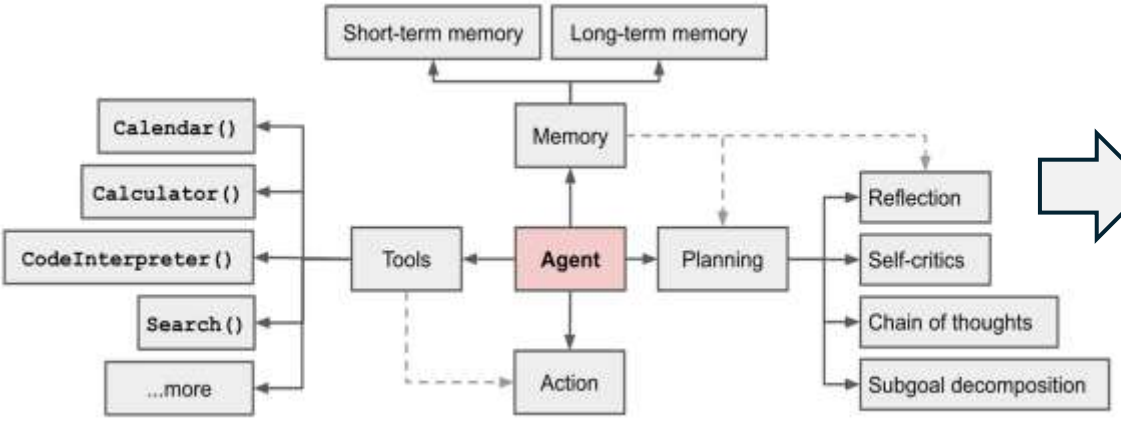


- **NMA is independent from the original controllers.**
  - A new east-west interface (**A2C**) needs to be added between NMA and controller to achieve capability calling and result feedback.
  - An intelligent southbound interface (“**A2N**”) needs to be added between NMA and the physical network.
  - NMA interacts with users (human or other agents) through a new **A2U** interface (**A2A interface** belongs to the scope of this interface)

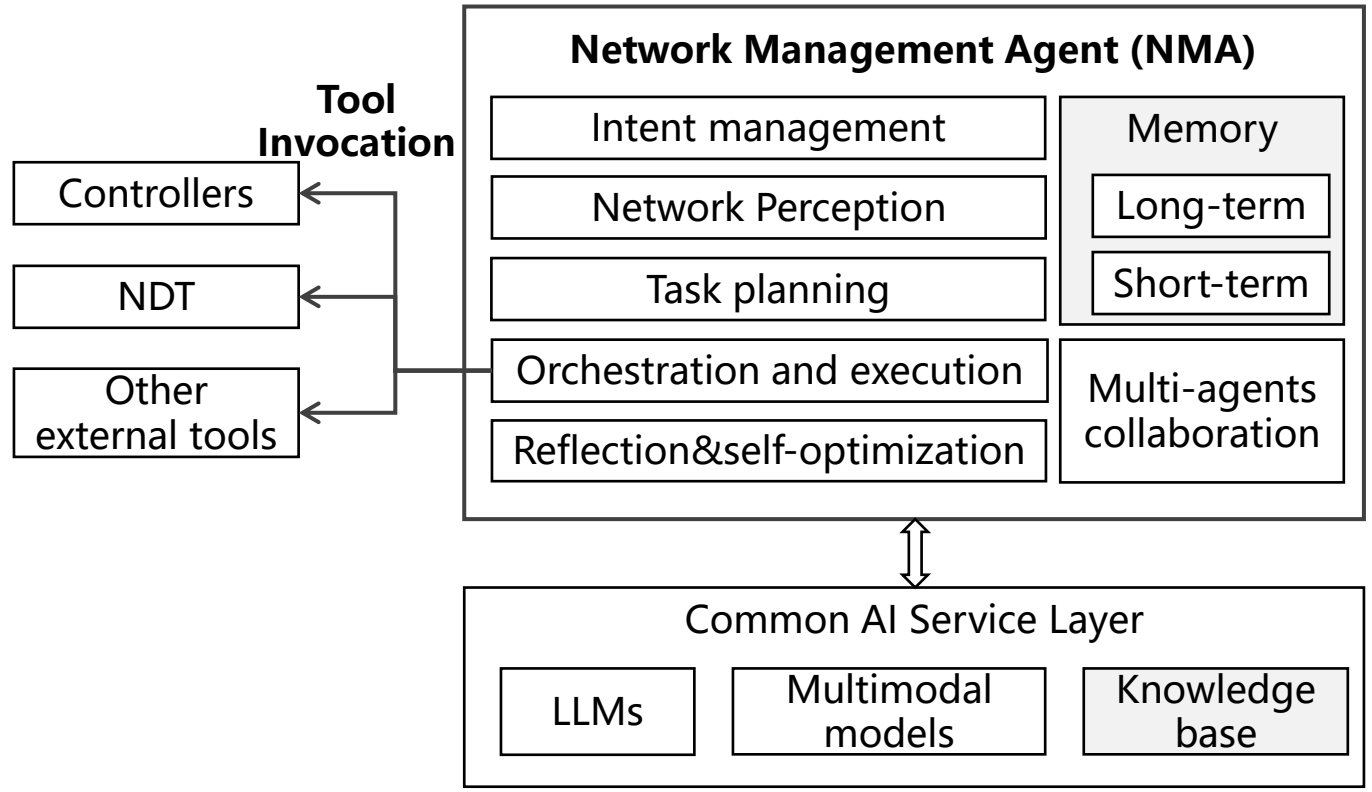
- **NMA is integrated and serves as a function of the controller**.
  - NMA interacts with original function modules through internal interface. NMA interacts with users through extended NBI (**including A2U**)
  - The enhanced MCS interacts with underlay physical network through extended SBI (“**E_SBI” including A2N**)

Major updates since 00-version (4)

By referring to the common AI agent framework, the draft updates **the reference function architecture of NMA**.



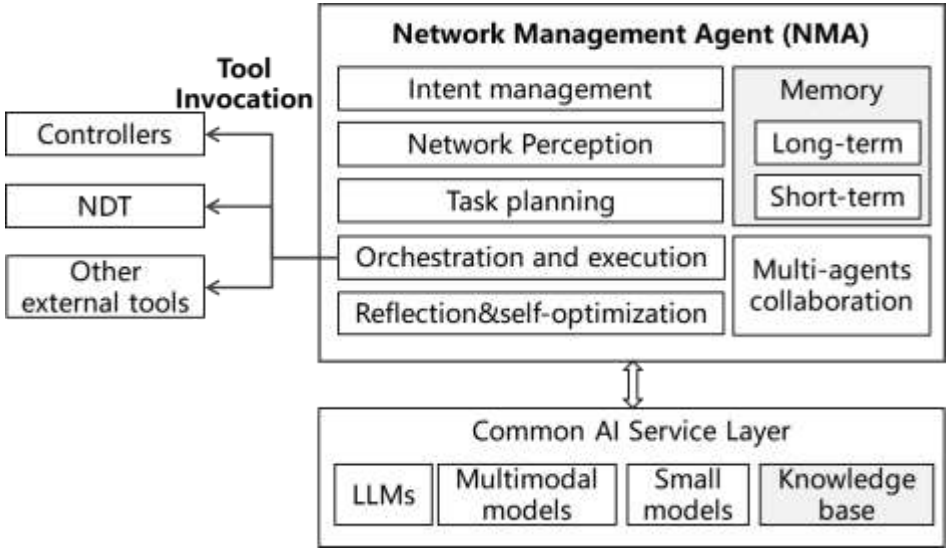
Common AI Agent framework



Reference network management Agent architecture

Call for feedbacks!

Function components of NMA

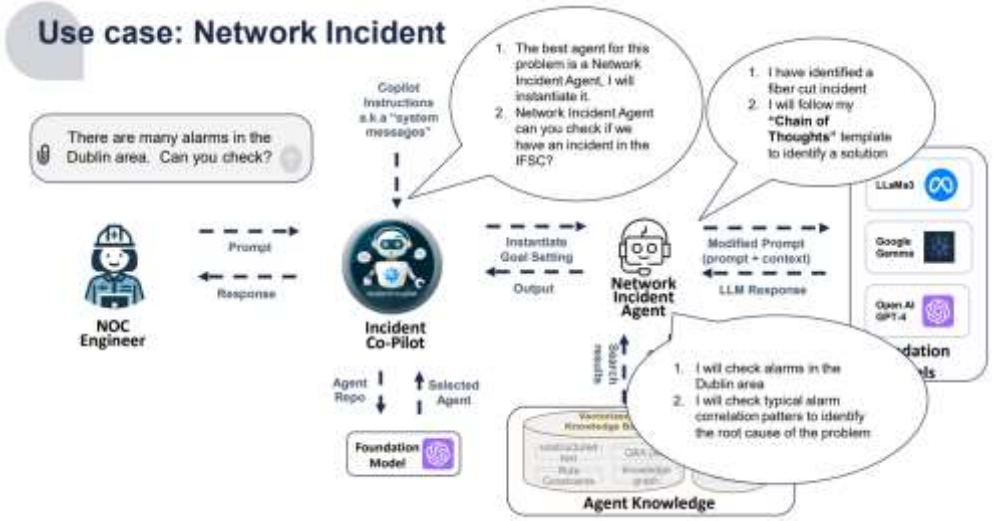


- **Intent Management:** Basic capability provided by AI models, responsible for collecting the input task and translate into intents through model reasoning.
- **Network Perception:** Achieve real-time query for network status information related to the task intent. Network status info are not limited to topology, service configurations, device status, alarms, performances, etc. The query source can be controller, NDT, etc.
- **Task Planning:** Based on the reasoning ability of AI models, break down the task intention into multiple sub operations.
- **Orchestration and execution:** Select the appropriate tools and automatically call the relevant tools or interfaces to perform the operation. After each sub operation is completed, the results of each operation are formed into task execution results.
- **Reflection and self-optimization**: Analyze and evaluate the historical decision-making process and execution results of the NMA. Based on the evaluation results, automatically adjust and optimize the NMA's strategies, parameters, etc. Artificial evaluation methods can be integrated to further optimize the NMA's performance through human supervision.
- **Memory:** Responsible for storing and processing data during the operation of NMA, including long-term memory (LTM) and short-term memory (STM).
- **Multi-agents collaboration**: Responsible for completing collaboration between multiple NMAs at different levels or in different application scenarios. The specific collaboration mechanism needs further research.

Function components of NMA

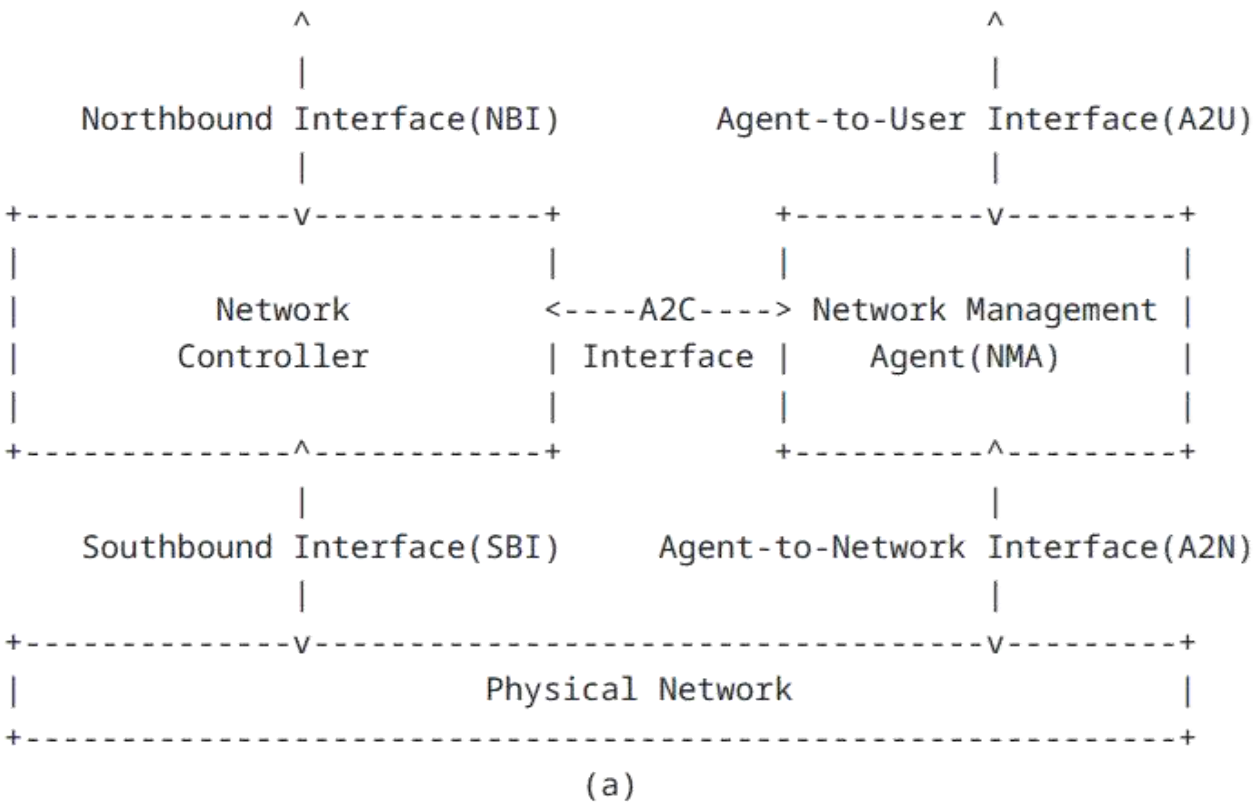
- **Common AI service layer:** including **various large language models (LLMs), multimodal models, small models, and knowledge base.** AI models provide public interactive intelligence capabilities as unified agent engine, to simplify NMA development. Knowledge base provides unified search for multi-type knowledge bases including vector knowledge base, system online help, operation and maintenance data logs), combines AI models to complete knowledge fusion and extraction, and improves the accuracy of NMA task execution.
- **Various NMAs can be constructed based on the common AI service layer.** During the operation of NMA, it leverages the model reasoning capabilities and knowledge base provided by the AI service layer to achieve functions such as intent parsing and task planning.
- Depending on the actual deployment requirements**, the AI basic service can also be deployed within the NMA.**

- For different application scenarios, there can be multiple scenario-oriented agents (like apps in the phone). **Aimed at the network planning, construction, maintenance, optimization, and operation scenarios, the main NMAs could include**:
  - Network Fault Handling Agent
  - Network Planning Agent
  - Network Optimization Agent
  - ……



Major updates since 00-version (5)

Interfaces of NMA



(a)

- **Agent-to-User interface (A2U):**
  - The interface between the NMA and the user, where the user can be a human or another system.
  - Used to receive call requests from users and return task processing results.
  - It should support both structured and natural language modes.
  - The **Agent-to-Agent (A2A) interface** between NMAs is included in the scope of this interface.
- **Agent-to-Controller interface (A2C):**
  - In the independent mode, this interface is an east-west interface between the controller and NMA; in the integrated mode, this interface is an internal interface of the controller and is not within the scope of this document.
- **Agent-to-Network interface (A2N):**
  - The interface between the NMA and the physical network.
  - In the independent mode, this interface is a new SBI between the Agent and the network; in the integrated mode, this interface can be extended based on the original SBI.

Major updates since 00-version (6)

Deployment modes considerations

- Due to an internal communication between the NMA and the controller, the independent deployment mode introduces several challenges to be analyzed, that can be grouped into “single agent” and “multi agent” challenges.

Single Agent

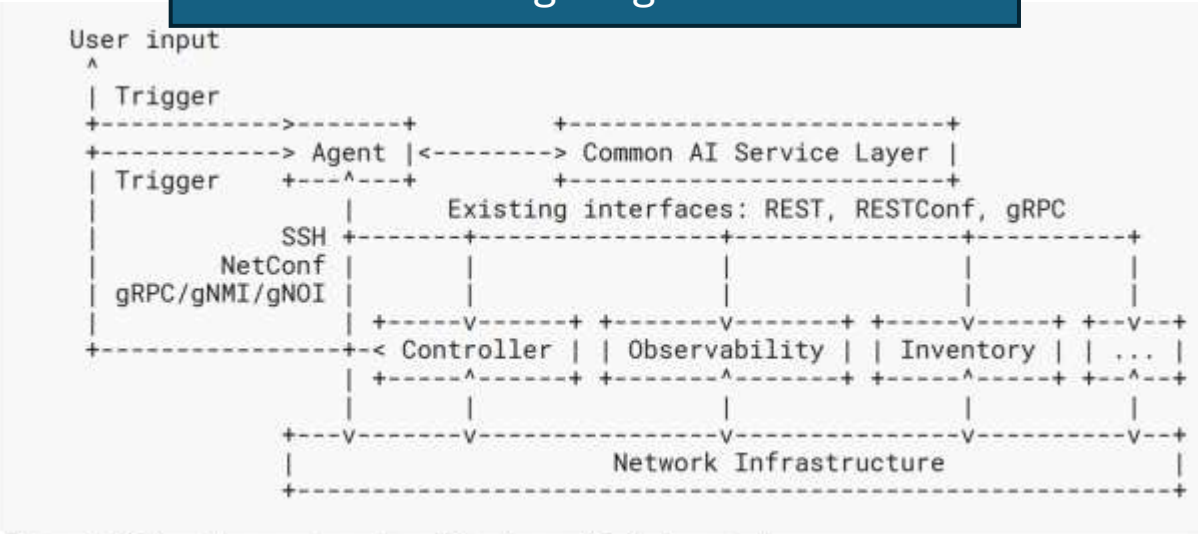


Figure 3: Network management architecture with single agent

- **Main issues to be addressed:**
- **NMA APIs:** A gap analysis against existing APIs needs to be carried out to understand if the NMA API requirements can be met and if we can find a way to describe network APIs for LLMs.
- **NMA triggers:** The trigger can be initiated by a controller or a human readable string. Natural language input or something with a more structured format?

Multi Agents

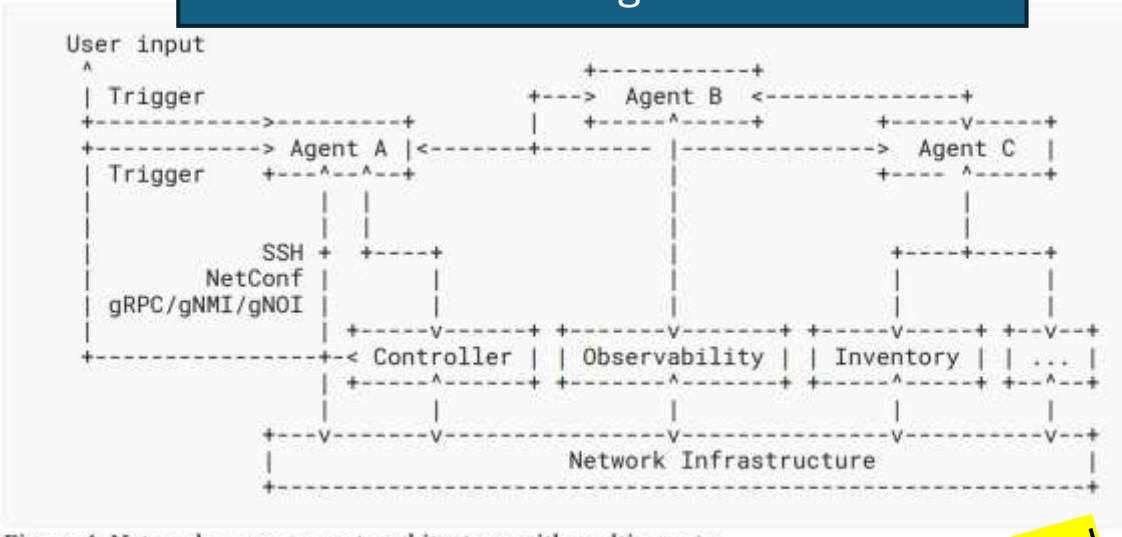


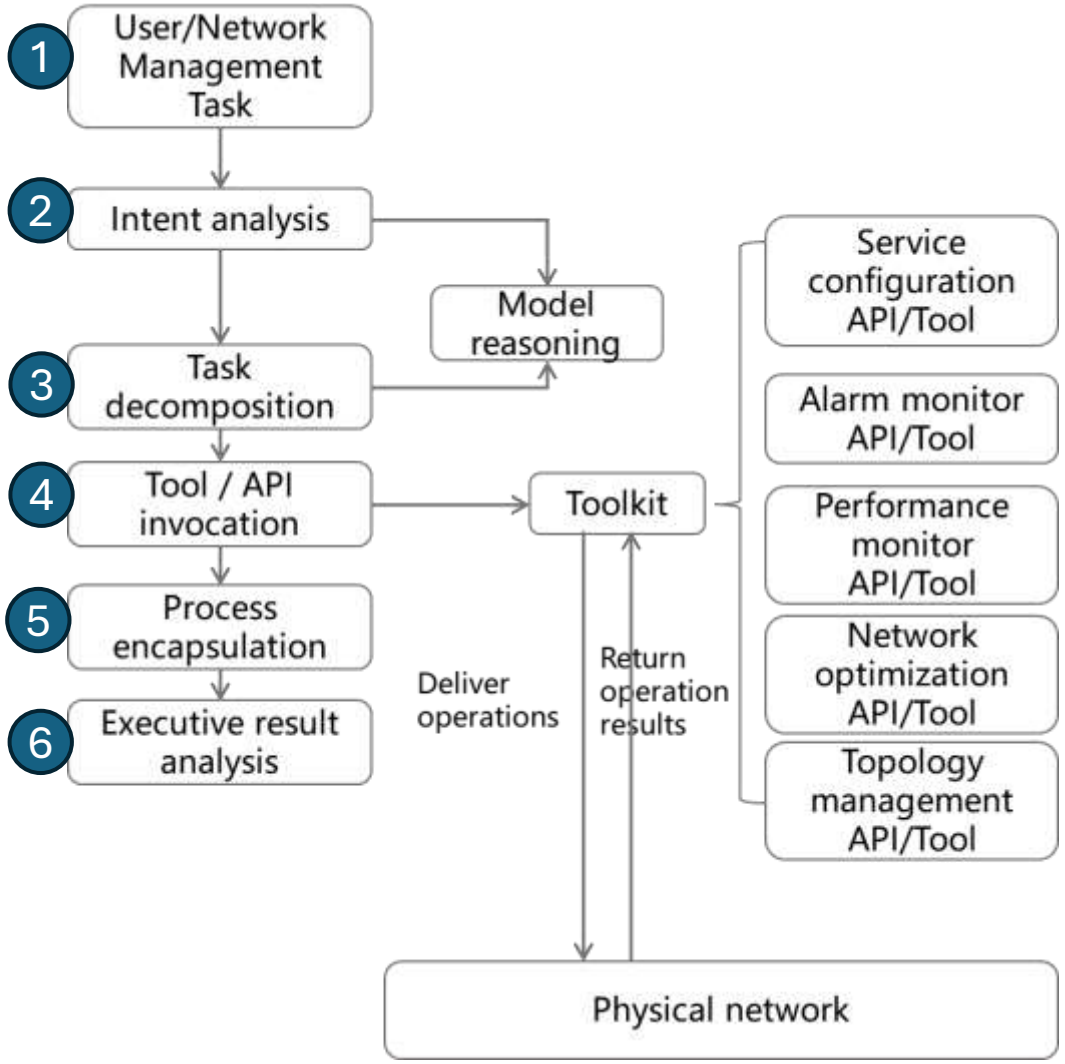
Figure 4: Network management architecture with multi agents

More complicated

Main issues to be addressed:

- **Inter NMA communication:** Necessary to support authentication.
- **NMA discovery:** NMA discovery and capability report mechanisms between multi-agents.

Common workflow of NMA (Unchanged)



① **User/Network Management Task Input:** Input the user's task information through multiple rounds of natural language interaction.

② **Intent Analysis**: Analysis user task intent through AI model reasoning provided by the AI based basic services within NMA.

③ **Task Decomposition**: **Split the task into detailed operations** to be performed based on the analyzed intent of the task.

④ **Tool/API Invocation**: Call the corresponding tool or function API to complete the execution of each operation listed in step 3). The ***toolkit*** refers to the collection of all tools that can be used directly to manage and operate physical networks.

⑤ **Process Encapsulation**: Package the operation results of all steps into the execution result of the entire task.

⑥ **Executive result analysis**: Analyze the task processing results and return to the user.

Goal: Achieve closed-loop automated processing of tasks and provide end-to-end intelligent network maintenance assistance.

Typical scenarios (Unchanged)

CAICT 中国信通院

① **Network management and maintenance scenarios, including:**

- **Intelligent planning and construction**: such as broadband installation, resource/capacity planning, intelligent acceptance, site selection, etc.
- **Intelligent maintenance**: such as intelligent fault diagnosis, quality analysis, operation and maintenance/cutting assistant, broadband maintenance assistant, etc.
- **Intelligent optimization**: such as route optimization, coverage optimization, topology optimization, and intelligent energy saving, etc.

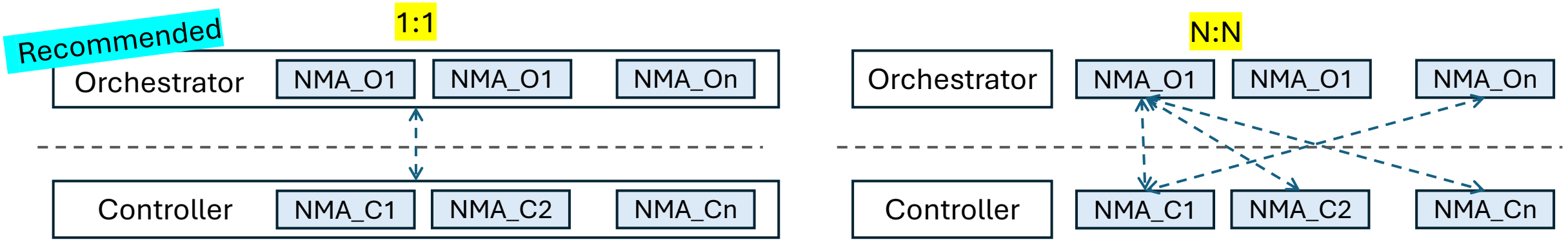
② **Network operation scenarios, including:**

- intelligent question and answer, customer service assistant, product recommend, intelligent marketing and other value-added services. *This part is outside the scope of this document.*

- ***Application considerations**: strong demand, feasible technology, and good input-output ratio, sufficient data for AI pre-training, perfect data annotations, high fault tolerance rate.*
- ***First possible mature application scenarios**: broadband installation and maintenance assistant, fault diagnosis, operation and maintenance assistant, etc.*

Open issues

- **Interfaces scheme of NMA: MCP, A2A? Are they suitable for Network Management Agents?**
  - *Glad to see that there are already related works:*
  - *Applicability of A2A to the Network Management (draft-yang-nmrg-a2a-nm-00)*
  - *Applicability of MCP for the Network Management (draft-yang-nmrg-mcp-nm-00)*
- **Multi-agent collaboration scenarios and mechanism: N:N or 1:1?**
  - **1:1:** Adopting the integrated mode enables one-to-one interaction between the controller and the orchestrator.
  - **N:N :** Adopting the independent mode results in many-to-many interaction relationships between multiple Agents at the Controller layer and multiple Agents at the Orchestrator layer, leading to complicated interaction.



Next Steps

- Improve the content of the draft, especially on the interfaces and workflows
- Collect comments and contributions
- Update the *03*-version at the next *IETF #124* meeting

Thanks for listening !

Any feedbacks are welcome!

zhaoxing@caict.ac.cn