Network Management
Internet-Draft
Intended status: Informational
Expires: 23 April 2026

Y. Hong
Daejeon University
J. Youn
DONG-EUI University
Q. Wu
Huawei
B. Claise
Everything OPS
20 October 2025

Motivations and Problem Statement of Agentic AI for network management
draft-hong-nmrg-agenticai-ps-00

Abstract

This document outlines the key objectives of introducing Agentic AI to the field of network management and highlights the fundamental issues with existing technologies that must be addressed to achieve these goals. It emphasizes the necessity for relevant groups within the IETF/IRTF and presents the core technological areas requiring standardization. The aim of Agentic AI is to facilitate a paradigm shift in which multiple autonomous AI agents collaborate to fully automate network operation, management and security.

Discussion Venues

This note is to be removed before publishing as an RFC.

Discussion of this document takes place on the Network Management Research Group mailing list (nmrg@irtf.org), which is archived at https://mailarchive.ietf.org/arch/browse/nmrg.

Source for this draft and an issue tracker can be found at https://github.com/billwuqin/agentic-ai-ps.

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 23 April 2026.

and restrictions with respect to this document.

Table of Contents

1. Introduction

The explosive growth of digital services and the increasing complexity of networks in 5G and future 6G environments demand real-time responsiveness, high efficiency and the ability to make autonomous decisions on a large scale from operational environments. To overcome the limitations of existing static automation methods and human-led Intent-Based Networking (IBN), a new Agentic AI-based paradigm is required. This involves introducing autonomous software entities that can interpret information, make decisions, perform meaningful autonomous actions and adjust plans in response to changing circumstances.

Unlike traditional automation, which relies on pre-programmed rules, agentic AI uses autonomous decision-making capabilities to handle large-scale network activities and customer requests swiftly and accurately. These agents perform tasks such as network traffic management, fault resolution, and customer interaction support, continuously executing responses that previously required manual human review or escalation.

Agentic AI uses large language models (LLMs) to encompass a wide variety of capabilities, such as reasoning, problem-solving, interacting with external environments and performing actions, which extend far beyond natural language processing. It can decompose tasks, breaking down complex objectives into specific tasks and

subtasks to achieve them. This cognitive capacity enables a persistent cognitive cycle (observation, inference, action), continuously aligning network operations with high-level business intent.

When such autonomous agents are widely deployed across the communications and network domains, standardized protocols are essential to ensure interoperability and security between different vendor platforms and network domains. The collaborative nature of agent-based AI systems (multi-agent systems, or MAS) means that standardized agent-to-agent protocols (A2A protocols) must be defined to prevent silos forming within the system and to facilitate discovery, understanding and collaboration between agents.

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.

3. Agentic AI for Network mMnagement

3.1. Role of Agentic AI in Network Operations

The complexity of network management and network operations are increasing exponentially, due to the increased size of networks and the increased frequency of change, for for the new 5G and future 6G services. This makes it increasingly difficult for existing automation techniques to meet the requirements for operational efficiency and service quality. Consequently, Agentic AI is an essential technological advancement for the realization of autonomous networks.

Agentic AI refers to intelligent systems that can act autonomously to achieve specific business objectives with minimal human supervision. These systems can reason through multi-step problems and adjust their actions in real time. Unlike passive traditional AI systems that respond only to direct commands, Agentic AI is an active system operating within an autonomous, closed-loop framework. This framework enables the system to perceive its environment, reason, plan a sequence of actions and execute them using various tools and APIs. This autonomy enables it to perform complex, multi-step processes such as software development, data analysis and network management.

The aim of autonomous networks is to leverage the capabilities of Agentic AI in order to transition operations and maintenance from static, human-managed, rule-based automation to dynamic, intent-based automation that is governed by humans. The ultimate goal is to reduce management costs and complexity, enabling rapid business optimization at unprecedented levels.

The primary objective of Agentic AI is to enable autonomous decision-making and the resolution of complex, multi-domain tasks. This is crucial in bringing operations closer to the level of autonomy that Agentic AI aims to achieve, by facilitating cross-domain collaboration. To achieve this, Agentic AI must align network capabilities with strategic business priorities, such as improving customer experience and reducing operational costs. This involves translating comprehensive business intent into localized, actionable network configuration plans.

Agentic AI optimizes resource allocation based on real-time demand and business objectives, enabling smarter resource and energy usage.

In architecture research for 6G, for example, the application of constrained agentic AI techniques focused on energy efficiency and secure real-time learning for dynamic resource allocation has been identified as a key objective [Agentic-AI-Wireless].

The Autonomic Networking Integrated Model and Approach (ANIMA) Working Group of the IETF developed the Autonomic Service Agent (ASA) for autonomic networking. [RFC7575] defines the ASA as An agent implemented on an autonomic node that implements an autonomic function, either in part (in the case of a distributed function) or whole [RFC7575]. In other words, the ASA is a core component of ANIMA: a software module that performs autonomic functions on network nodes. The ANIMA Working Group is defining design guidelines, lifecycle management, authorization and coordination standards for the ASA [ANIMA].

IETF' AI Preferences (AIPREF) Working Group is focused on standardizing a common vocabulary and mechanism through which users and systems can express their preferences regarding the use of their content in the development, training, deployment and use of AI models [AIPREF].

3.2. Operation of Agentic AI for Network Management

The principal components of agentic AI can be broadly divided into the intelligence core and the execution tool domain.

3.2.1. Intelligence core

The intelligence core is responsible for an agent's decision-making and problem-solving capabilities. Large language models (LLMs) or specialized AI models form the basis of this core. Reasoning Engine/ LLM: This constitutes the core of the agent's brain. It understands abstract objectives (intent) received from users or higher-level systems, creates step-by-step plans (plan) to achieve them, evaluates the outcomes of execution (reflection) and uses logical reasoning to modify plans or determine subsequent steps.

Memory is the data repository that agents learn from and refer to.

* Short-term memory: It stores the context of the current task and recent execution results.

* Long-term memory: It stores persistent information such as previously successful solutions, general knowledge and network architecture guidelines.

The tool Orchestrator manages the list of external tools (APIs, functions) available for agents to use. During the planning phase, it determines which tool is most appropriate and, during the execution phase, it is responsible for calling the tool and accurately configuring the necessary parameters.

3.2.2. Execution & Interaction

These components enable the agent to communicate with and make changes to the external environment (i.e. the network or system).

* Tool set/capability: A collection of all the external interfaces that an agent uses to perform tasks within a network environment.

* Execution environment: A sandbox environment in which code generated according to the plan is executed safely, and external tools are invoked.

* Sensing/observation mechanism: The channel through which the agent

verifies execution results and collects the current environmental state. This involves more than just invoking tools; it continuously draws network events, sensor data and similar inputs into a feedback loop.

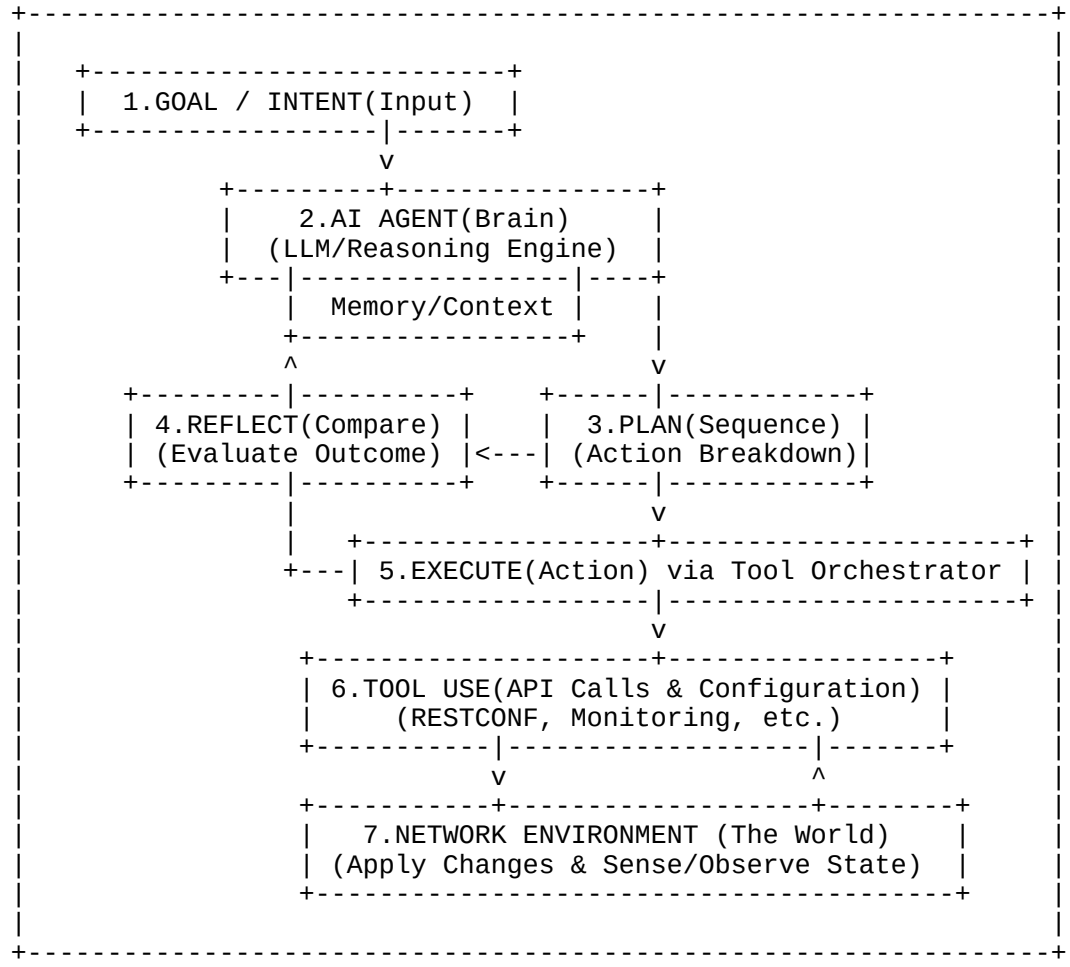


Figure 1: Execution & Interaction

4. Problem Statement of Existing Techniques for Agentic AI

4.1. Architectural Bottlenecks and the Failure of Centralization

Existing AI and automation systems have often relied on centralized infrastructure for data aggregation and heavy computing. However, these centralized models cannot handle the volume, velocity, and distributed nature of Agentic AI workloads. Centralized AI systems are constrained by central infrastructure, resulting in high latency due to round-trip times to the cloud. Such latency is unacceptable for real-time applications such as self-healing and 5G slicing management. There is also the issue that the central server becomes a bottleneck, limiting scalability. The inherent limitations of such centralized models (single point of failure (SPoF), latency) inevitably drive Agentic AI architectures towards a distributed mesh form. This leverages local processing at the edge for low latency and fault tolerance, requiring the standardization of distributed control and communication mechanisms that transcend conventional centralized SDN/management models.

4.2. Absence of agent-to-agent (A2A) Semantic Interoperability

Agentic systems are often built by different vendors using various frameworks, leading to fragmented and siloed system operations. Complex network management tasks require the decomposition of work

and collaboration between specialized agents. Without standardized agent-to-agent (A2A) protocols, bespoke connectors become necessary to connect these fragmented systems, slowing down development and integration speeds.

Standardization must define consistent payloads and interfaces that support real-time interactions between systems, enabling agents to discover, understand, and collaborate with one another regardless of their underlying implementations.

4.3. Lack of Dynamic Trust and Accountability in Autonomous Behavior

The introduction of AI agents as autonomous entities performing actions at machine speed presents significant security and governance challenges. Traditional identity and access management (IAM) focuses on human users or predefined roles. However, autonomous agents operate with dynamic intent, require context-aware access, and must maintain provable accountability for every action they perform. Without a robust Zero Trust framework specifically designed for non-human autonomous entities, there is a risk of catastrophic security breaches or manipulation where autonomous systems could outpace human control capabilities.

4.4. Real-time Data Validity and Resilience Issues

The decision-making of AI agents is determined by the quality of the data they receive. In a network environment, data quality is of paramount importance. Incomplete, delayed, semantic-less, context-less, or corrupted data feeds can lead to severe operational or financial losses when agents take autonomous actions (e.g., traffic rerouting, forced execution of financial transactions). Therefore, it must extend beyond the current focus on bandwidth and speed to include quality verification of the data agents rely upon and resilience of service paths. This is essential to meet the requirements of continuously operating intelligent agents.

4.5. Problems with the Existing IBN System: Rigidity of the Intent Translation Engine (ITE)

Existing IBN systems rely on the Intent Translation Engine (ITE) or the Intent-Based System (IBS) spatial functionality to bridge the gap between the business intent and the network operational infrastructure. This translation is typically driven by predefined data models such as YANG models and lacks the necessary adaptive flexibility when unforeseen conditions arise. IBN fundamentally shifts operational modes to a dynamic intent-based approach, yet retains the inherent limitation that control remains under human oversight. Agentic AI minimises or eliminates human intervention in this cognitive loop through LLM-based reasoning and planning capabilities, refining the IBN closed loop by integrating continuous reasoning and conflict resolution capabilities into the cognitive layer. These capabilities represent what was lacking in the classical IBN definition and form the core technical objective.

4.6. ANIMA ASA’s Problem: Cognitive Simplicity

ANIMA’s ASAs are typically designed for specific, localized autonomous functions (e.g., prefix management, bootstrapping). They rely heavily on predefined policy structures and lack the complex reasoning, planning, or self-reflection capabilities characteristic of Agentic AI (LLM-based task decomposition). ANIMA’s ASA is conceptually a precursor to Agentic AI, but lacks a cognitive core (LLM/inference engine). Agentic AI introduces LLM-based planning and tool-use capabilities that require complex, semantic negotiation (A2A) beyond simple information exchange (GeneRic Autonomic Signaling Protocol; GRASP), demonstrating the necessity for a dedicated

protocol layer that extends beyond the existing ANIMA framework.

5. Objectives of Agentic AI for Operations & Management

5.1. Objective 1 - Autonomous Network Operations & Management

Beyond minimizing human intervention, it must implement a Autonomous Driving Netowrk (defined in TMF) that autonomously recognises, diagnoses, infers, and resolves issues even in unpredictable situations.

Key Features:

* Predictive & Proactive Fault Management: AI agents learn traffic patterns, logs, and performance metrics in real time to identify potential causes before faults occur. The network autonomously reroutes traffic or reallocates resources to prevent service interruptions at source.

* Intelligent Root Cause Analysis: In complex, intertwined fault scenarios, multiple agents collaborate to synthesize distributed data. They deduce the root cause as a "problem of correlations" rather than a single point of failure and propose solutions.

* Autonomous Configuration & Optimization: AI agents comprehend high-level objectives such as ‘optimize user experience’ and autonomously configure and continuously fine-tune routing protocols, QoS policies, security rules, and other elements to achieve them.

5.2. Objective 2 - Intelligent & Dynamic Resource Orchestration

To address unpredictable traffic demands such as 6G, holographic communications, and large-scale IoT, network resources (computing, storage, bandwidth) are allocated and coordinated in real time and proactively.

Key Features:

* Dynamic Network Slicing: AI agents recognize application requirements (latency, bandwidth, etc.) in real time, instantly creating, scaling, and downsizing customized network slices per user or service.

* Cross-Domain Resource Negotiation: AI agents distributed across networks of different telecommunications or cloud providers negotiate in real time to dynamically secure optimal resources, ensuring end-to-end quality for global services.

* Edge Computing Resource Optimization: By predicting edge node load and user mobility, AI agents dynamically reallocate workloads to optimal edge nodes while ensuring service continuity.

5.3. Objective 3 - Predictive & Adaptive Network Security

Beyond defending against known attack patterns, AI agents autonomously detect unknown zero-day attacks or advanced persistent threats (APTs) and reconfigure defence systems in real time.

Key Features:

* Autonomous threat hunting and response: Security agents continuously detect minute anomalies across the entire network. If an anomaly is deemed a threat, they respond immediately by taking action such as isolating infected nodes or blocking attack traffic, all without human intervention.

* Dynamic Defense Posture: AI agents dynamically modify firewall policies, access control lists (ACLs), and traffic filtering rules in real time based on attack type and intensity, thereby minimizing the attack surface.

5.4. Objective 4 - Enabling Novel Network Service Models

By transforming the network itself into a single, vast distributed AI platform, it enables new communication services and business models that were previously impossible.

Key Features:

* Intent-driven Service Creation: When a user requests in natural language, 'I want to play a lag-free VR game with my friends,' an AI agent interprets this and provides a Network-as-a-Service that instantly allocates the necessary resources (such as network slices and edge servers).

* Semantic Communication: Communication focuses on the "meaning" or "purpose" conveyed by data rather than the bits themselves, enabling ultra-efficient communication that achieves maximum effect with minimal data transmission.

5.5. Objective 5 - Autonomous, High-Fidelity & Action-Aware Network Measurement

To turn raw network telemetry into trustworthy, context-rich insight that continuously retrains itself, explains its own uncertainty, and feeds closed-loop control without human analysts.

Key Features: - Generative Telemetry Synthesis & Gap-Filling: Gen-AI models learn multi-modal telemetry (packets, flow records, SNMP, syslogs, DPI, spectrum scans) and hallucinate statistically faithful "missing data" where sensors are sparse or silent, delivering 100 % coverage at any time/space scale.

* Semantic Anomaly Narratives & Root-Cause Metrics: Instead of threshold alerts, the model outputs human-readable stories ("Between 02:13-02:19 UTC, TCP RTT on slice-C rose 38 % because 17 % of ECN-marked packets were re-routed via the Seattle POP due to a mis-announced BGP community"). Each sentence is back-traced to verifiable measurement samples.

* Self-Driving Measurement Campaigns: The AI translates high-level intents ("tell me if user-perceived 4 K latency could exceed 150 ms during the next football final") into dynamic sampler schedules, probe paths, and packet structures; it launches the campaign, stops when statistical confidence is reached, and releases resources back to the data plane.

* Counterfactual & Predictive "What-if" Metrics: Given a proposed config change (new AQM, additional slice, 400 GbE upgrade), the generator produces the expected delay/loss/jitter distributions before any byte is moved, letting operators compare KPI deltas without real-world probing.

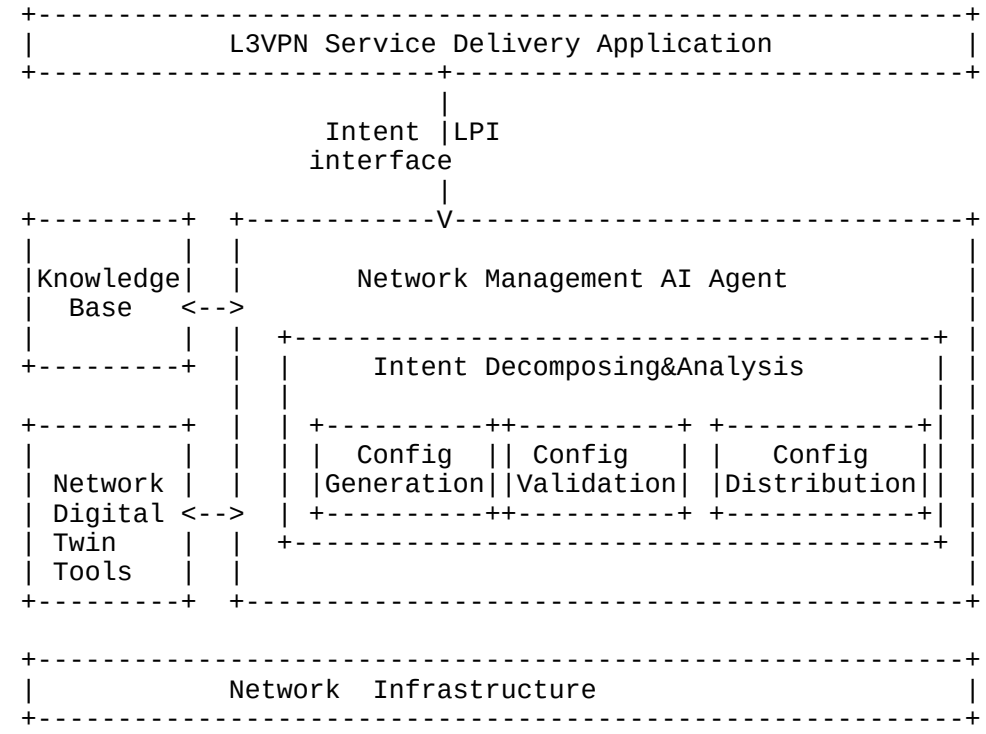
6. Use cases of Agentic AI for Operations & Management

Different use cases for Agentic AI on Operation & Management can be identified, as described in the following sections.

6.1. Intent Based Service Delivery

Below is the diagram showcasing how network management AI agent takes

effect on the intent based service delivery process.



Legend: LPI - Language Programming Interface

Figure 2: Intent Based Service Delivery

Step a. L3VPN Service Delivery Application at the OSS layer uses Language Programming Interface (LPI) to send service intent request "Create L3VPN service with 2 VPN sites in London and Paris using L3SM Service Model".

Step b. The Network Management AI Agent looks up knowledge base to understand the intent and identify user's objective "VPN Service Creation".

Step c. The Network Management AI Agent further interacts with Knowledge base for expert experience and looks up thought of chain related to "VPN Service Creation". And then the Knowledge base returns results to the Network management AI Agent.

Step d. The Network Management AI Agent decomposes user intent and break down the tasks into operational workflow including configuration generation, configuration validation, configuration distribution. For configuration validation, it will interact with Network Digital Twin tools to obtain the validation results.

Step e. After L3VPN Service is delivered successfully, the Network Management AI Agent will use LPI to return success results.

6.2. Cross-layer and Cross-domain Multi-Agent communication for Complaint handling

In this scenario, automotive companies centrally collect complaints from their customers (drivers) and use the operator' s complaint system to feedback issues to the operator. The operator's BSS trouble ticket system generates tickets from these complaints and dispatches them to the OSS. The integrated vehicle networking complaint handling agent within the OSS analyzes the trouble tickets and performs fault localization. The ticket will be sent to the corresponding vehicle networking trouble ticket agent within OSS

based on whether fault localization is within or beyond specific maintenance domain.

The vehicle networking trouble ticket agent within the OSS will parse the ticket into multiple multi-steps workflow and interact with the IP network agent and mobile network agent within its management domain to resolve the problem.

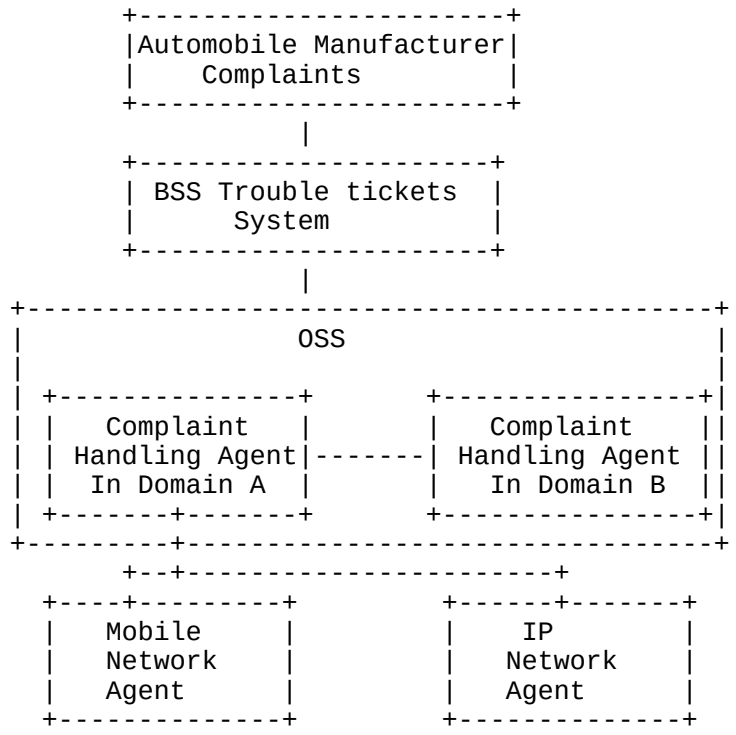


Figure 3: IoV User Complaints Handling

o Tasks are triggered by natural language

* Complaints usually come from end-users or enterprises

  - who may not have a deep understanding of network

    o sometimes are unable to provide accurate descriptions

o Tasks possess both abstraction and expertise

* Abstraction: complaint content is unpredictable and the involved domains cannot be anticipated

* Expertise: The final closed-loop of the task depends on the network

o Tasks involve cross-layer and cross-domain aspects

* Cross-Layer: BSS/OSS -> Network

* Cross-domain:

  - Technical domains (wireless network domain, backhaul network domain)

  - management & maintenance domains (i.e. across provinces and cities)

6.3. AI Agent Driven Network Management

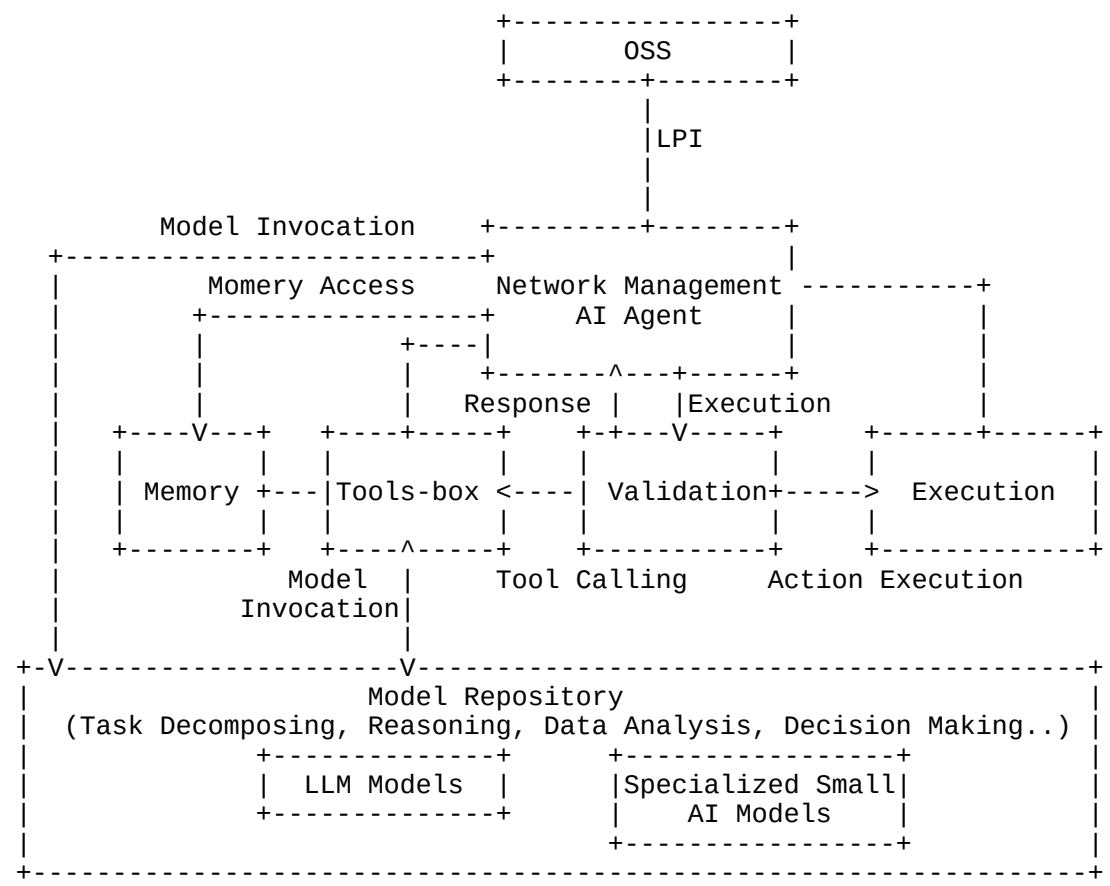


Figure 4: AI Agent Driven Network Management

Traditional network operation and maintenance require extensive human oversight and are constrained by predefined policies, limiting real-time adaptability. Network management AI agents at the network level enhance network intelligence and automation by integrating large network foundation models, specialized small AI models, and feedback closed loops mechanisms. The key functional requirements of the Network management AI agent include:

* Integrate with large foundation models and specialized small models for context-aware decision-making;

* Support Intent realizing including task decomposition,reasoning, inference&prediction and decision making.

* Support autonomous execution of network service lifecycle management, including network service delivery, network anomaly detection, predictive maintenance and troubleshooting, network re-optimization;

* Work with upper layer OSS to facilitate cross-layer collaboration, enabling seamless communication between network elements;

7. Security Considerations

TODO Security

8. IANA Considerations

This document has no IANA actions.

9. References

9.1. 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/rfc/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/rfc/rfc8174>.

9.2. Informative References

[Agentic-AI-Wireless] "Advanced Architectures Integrated with Agentic AI for Next-Generation Wireless Networks", 2025, <https://arxiv.org/html/2502.01089v3>.

[AIPREF] "IETF AIPREF WG", 2025, <https://datatracker.ietf.org/group/aipref/about/>.

[ANIMA] "IETF ANIMA WG", 2025, <https://datatracker.ietf.org/group/anima/about/>.

[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/rfc/rfc7575>.

Acknowledgments

TBA

Authors' Addresses

Yong-Geun Hong
Daejeon University
62 Daehak-ro, Dong-gu
Daejeon
34520
South Korea
Email: yonggeun.hong@gmail.com

Joo-Sang Youn
DONG-EUI University
176 Eomgwangno Busan_jin_gu
Busan
614-714
South Korea
Email: joosang.youn@gmail.com

Qin Wu
Huawei
101 Software Avenue, Yuhua District
Jiangsu
210012
China
Email: bill.wu@huawei.com

Benoit Claise
Everything OPS
Belgium
Email: benoit@everything-ops.net