CATS Reference Model for AI-Agent Communication Network (ACN)

draft-jiang-cats-reference-acn-00

Tianji Jiang, Peng Liu

Nov. 01 ~ 07, 2025, IETF 124 @ Montreal, Canada

Agenda

- AI-agent and ACN (AI-agent Communication Network)
- ACN realization: 3 network communication models
- CATS-optimization for AI-agent services
- CATS-like metrics model for ACN
- AI-agent protocol stack for CATS metrics extension
- CATS reference model, reference points & interfaces for ACN

AI-agent and ACN (AI-agent Communication Network)

AI agents:

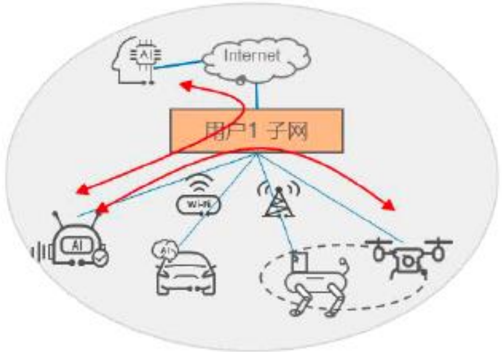
- SW-driven entities with embedded AI, including *ML, NLP*, and more.
- Instantiation: either in the physical state as embedded HW devices (e.g., robots) or in the virtual state (e.g., SW-implemented Apps)
- Crucial in network domain (particularly in Mobile space): enhancing network efficiency, predicting network conditions, optimizing resource consumption, making autonomous & intelligent decisions, and facilitating seamless communication among servicing & serviced entities

AI-agent Communication Network (ACN):

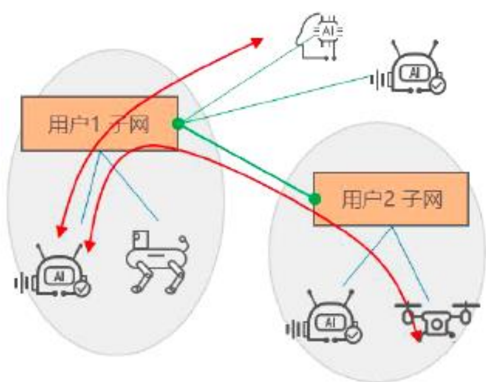
- An interconnected network framework tailored to advance the communication among AI-agents. It could reach the global coverage.
- Satisfy the on-demand communications, interactions & collaborations with secure and controllable information flow paths for AI-agents in distributed deployment mode.

ACN Realization: 3 Network Communication Models

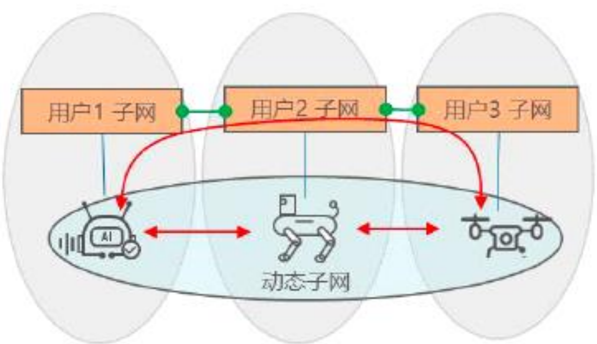
- **<u>Static</u> <u>intra</u>-domain only ACN**: the communication model for AI-agents residing within a *single administrative domain or network*. AI-agents in the domain form a network group maintaining the static communication association. AI-agents inside the domain do not communicate with AI-agents outside the domain.



- **<u>Static</u> <u>inter</u>-domain ACN**: the communication model for AI-agents that reside *across multiple administrative domains*. AI-agents across domains form one or more network groups maintaining the static communication associations. AI-agents in a domain can communicate with AI-agents both inside and outside the domain.



- **<u>Dynamic</u> <u>multi</u>-domain ACN**: the communication model for AI-agents that may dynamically form a network group to handle a temporarily-generated task. These AI-agents could be in the same or different administrative domains. This is a dynamic communication mode versus the previous two static modes.



**Pictures from the CMCC White Paper "AI-agent Communication Network", July 2024.*

CATS-optimization for AI-agent Services

- Dramatic variations among AI demands, data processing tasks and capabilities of AI agents.
  - Case#1: the compute power of lightweight terminals, e.g., smartphones, is difficult to handle locally the complex computation tasks with more than billions of parameters.
  - Case#2: the challenge of real-time responsiveness if complex tasks are delegated to more powerful (AI-agent) instances deployed remotely in clouds
  - More cases (from 3GPP): Ambient IoT, Integrated sensing, …, etc.

CATS-like metrics model for ACN:

- (Existing) Network metrics: For network entities like routers or switches, they can be bandwidth, capacity, throughput, transmission delay, TX bytes, RX bytes, host bus utilization, etc.
- (Existing) Compute metrics: For compute nodes, end servers, and/or service instances, they can be CPU, GPU, NPU, memory, storage, system delay.
- (New) AI-agent metrics: AI-agent functionalities & capabilities, AI model types, #parameters of models, authentication/authorization policy, etc.

AI-agent Protocol Stack for CATS Metrics Extension

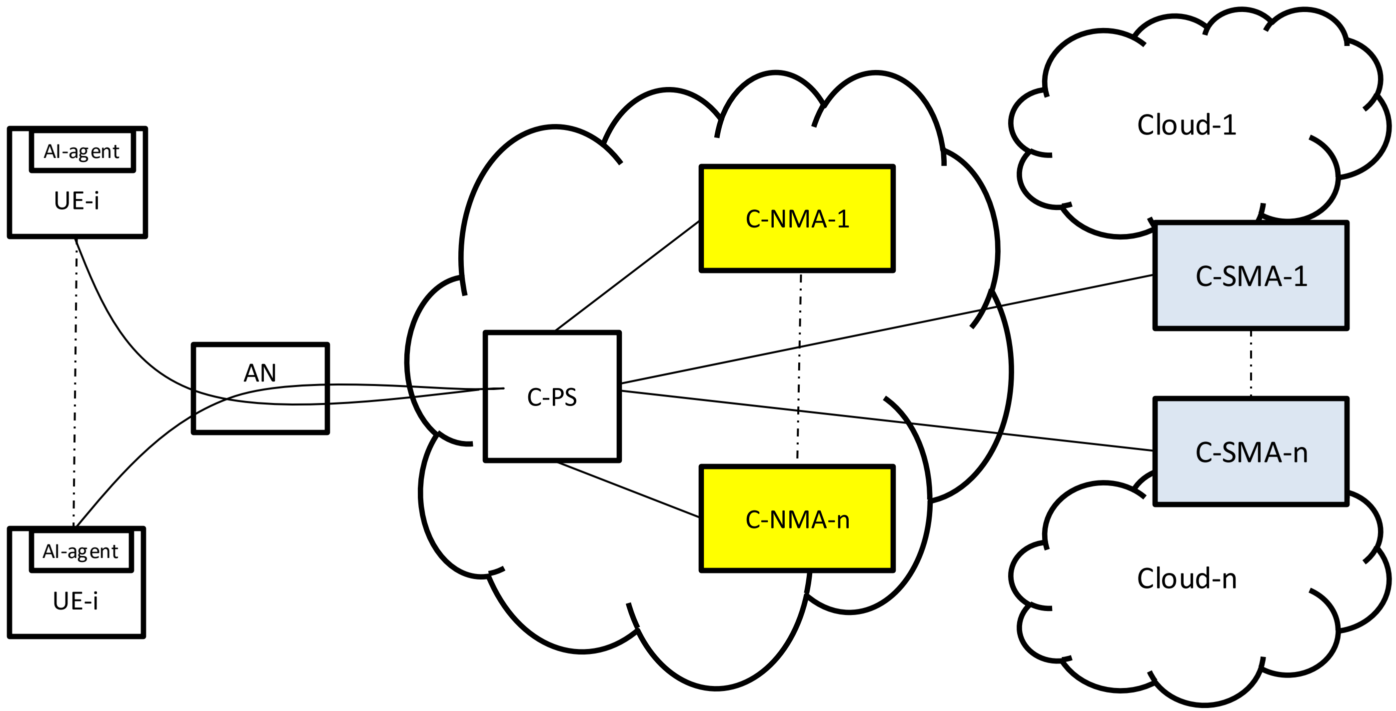
- (Existing) CATS metrics (network & compute):
  - Targeted toward the traffic steering at the network layer, which might be achieved via the network protocol extension
- (New) AI-agent metrics:
  - Generated for the overlay-exchange among App clients (those served as AI-agents), which are not generally subject to the signaling path over network protocols.
  - AI-agents in an ACN reside above the network & transport layers (most likely on App-layer)

```
===> Extended CATS for ACN
        +-----------------+    AI-agent Metrics +-----------------+
        |     AI Agents   |<------------------->|     AI Agents   |
        +-----------------+                     +-----------------+

===> Existing CATS here
        +-----------------+    Compute Metrics  +-----------------+
        | Compute-Entities|<------------------->| Compute-Entities|
        +-----------------+                     +-----------------+

        +-----------------+   Network Metrics   +-----------------+
        | Network Entities|<------------------->| Network Entities|
        +-----------------+                     +-----------------+
```

CATS Reference Model, Reference Points & Interfaces for ACN (1)

To define reference model, Ref. Interfaces & Ref. points:

- Existing CATS: C-PS/C-NMA, C-PS/C-SMA, C-NMA/C-NMA, C-SMA/C-SMA
- Extension w/ ACN: AI-agent/C-PS, AI-agent/AI-agent, '+ Existing'



CATS Reference Model, Reference Points & Interfaces for ACN (2)

- **Service model**: Proposed to be a producer-consumer model.
- **Reference interfaces and points** (between entities): definition, identity, name, parameters, etc.
- **Signaling messages**: definitions, parameters, types of exchanged messages (network-, compute-, and AI-agent metrics), etc.
- **Signaling & management procedures**: CATS entity registration, authentication and authorization; peer discovery and selection, peering session establishment, message-exchange, peering-state sync-up, … , etc.
- **Overlay-based** implementation and protocol stacks.
- **Communication channels & implementation** schemes: authenticated APIs, message brokers.

```
+---------------+                      +-------------+
|   CATS C-PS   +(c1)------//------(c2)+ CATS C-NMA  |
+---------------+      RP_ps_nma       +-------------+

+---------------+                      +-------------+
|  CATS C-NMA   +(c1)------//------(c2)+ CATS C-NMA  |
+---------------+      RP_nma_nma      +-------------+

+---------------+                      +-------------+
|    AI-agent   +(c1)------//------(c2)+ CATS C-PS   |
+---------------+      RP_aia_ps       +-------------+
```

Figure 3: Reference points & interfaces btwn C-PS and C-NMA

Summary

- AI-agent, ACN, and ACN realization/communication models
- CATS-optimization for AI-agent services, CATS-like metrics model for ACN
- AI-agent protocol stack for CATS metrics extension
- CATS reference model, reference points & interfaces for ACN

Is the CATS WG interested in defining *a reference model for CATS*? If yes, using this document as the baseline?

Thank you!