Computing-Aware Traffic Steering (CATS) Operations, Administration, and Maintenance (OAM) Framework

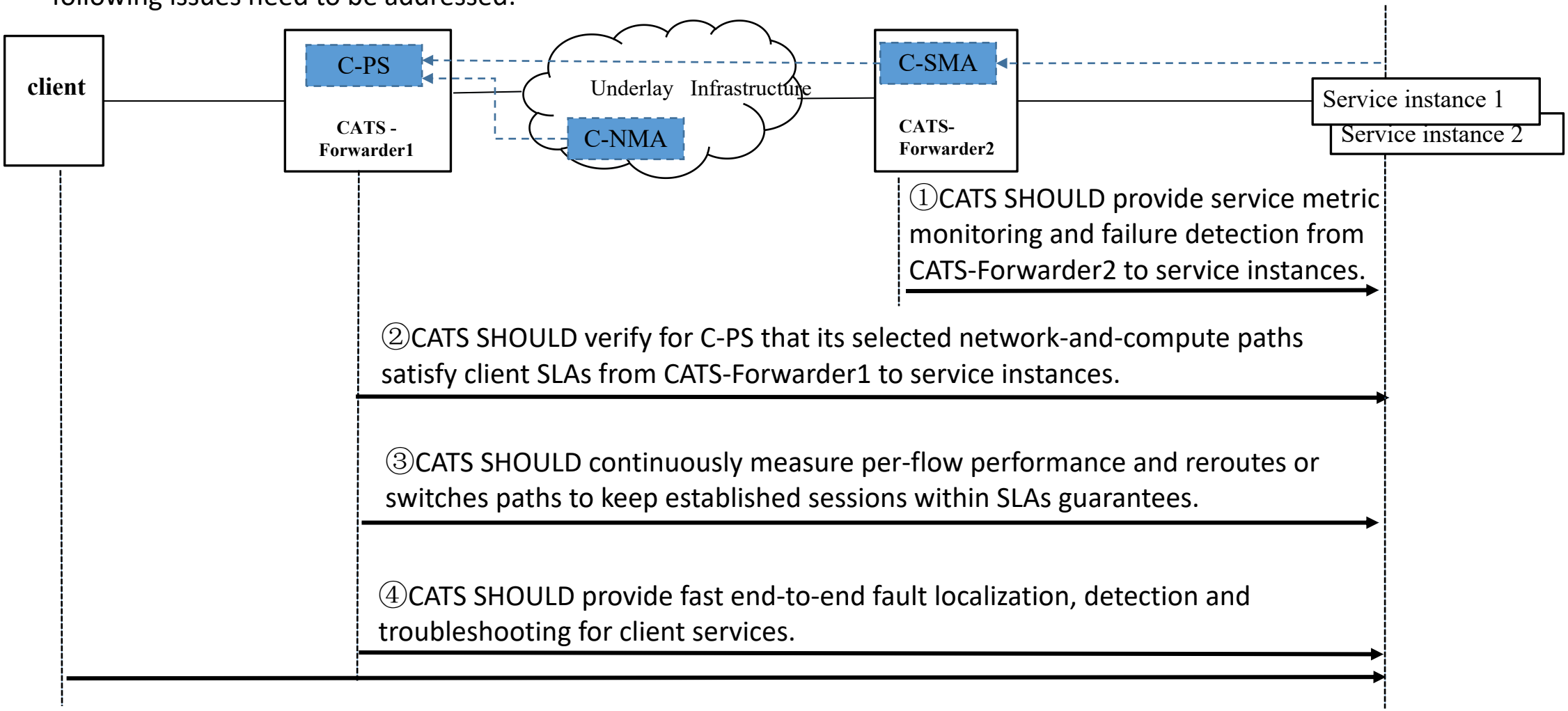
draft-fu-cats-oam-fw-04

Huakai Fu (ZTE Corporation)
Bo Liu (China Mobile)
Zhenqiang Li (China Mobile)
Presenter : Quan Xiong (ZTE Corporation)

IETF-124, Nov. 2025

Motivation

- As per draft-ietf-cats-framework, the C-PS collects service metrics from C-SMA and network metrics from C-NMA to select the Egress CATS-Forwarder for accessing a service-contact instance. To provide adequate OAM for CATS, the following issues need to be addressed:



CATS OAM Layering Model

- **Link OAM**: failure detection and performance monitoring are conducted for a single link. **In CATS OAM, the link OAM should cover the link from CATS-Forwarder2 to service instances.**

- **Tunnel OAM**: a tunnel bears multiple services and failure detection and performance monitoring are conducted on the LSP layer. **In CATS OAM, the tunnel OAM covers the network paths from CATS-Forwarder 1 to CATS-Forwarder 2.**

- **Service OAM**: conducted on the service layer that is provided by the network to evaluate the service quality. **In CATS OAM, the service OAM should cover the service instances.**

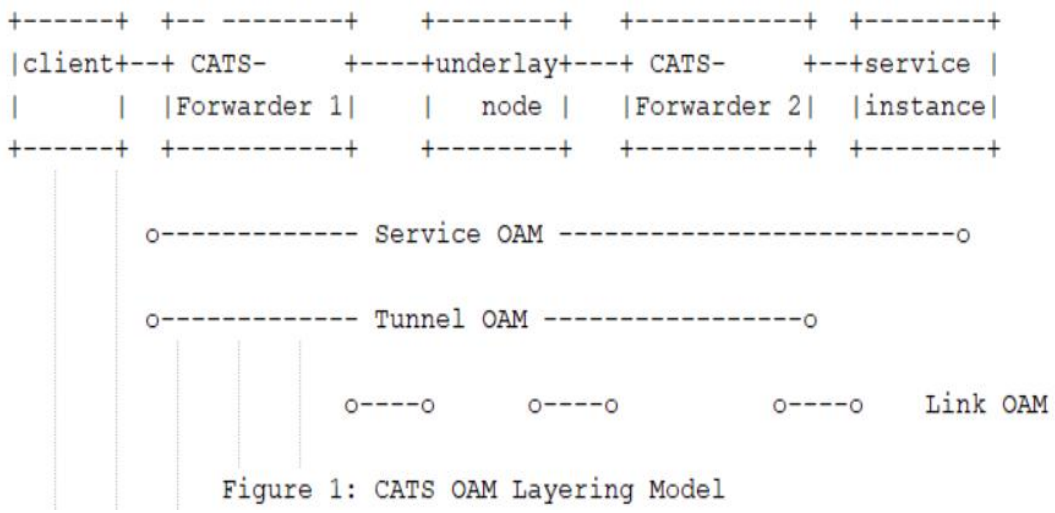


Figure 1: CATS OAM Layering Model

CATS OAM Functional Components

- **SI-OAM (Service Instance OAM)**
  - Detects failures and measures the service metrics between egress CATS-Forwarder and service instances.
  - Feedback to C-SMA for quick failure detection.
- **TC-OAM (Traffic Classifier OAM)**
  - Monitors the paths from the ingress CATS-Forwarder to service instances to ensure SLA through metrics monitoring.
  - Triggers recalculations if SLAs not met.
- **AF-OAM (Active Flow OAM)**
  - Assesses and maintains service performance in CATS.
  - Adjusts paths dynamically for user experience.
  - Prompt troubleshooting across the network.

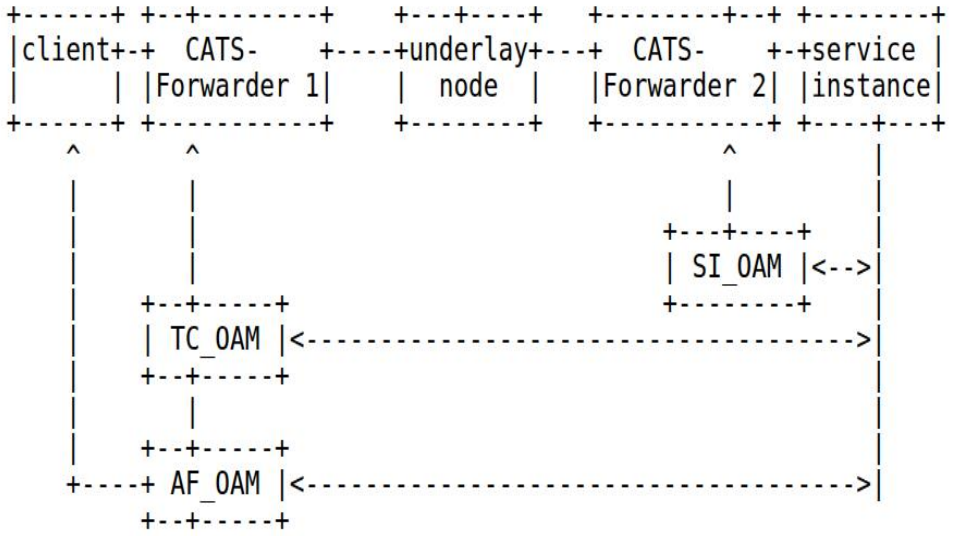


Figure 2: CATS OAM Functional Components

CATS OAM Requirements

Types	Requirements
Operation	* **Sub-second telemetry on CPU/GPU/memory/accelerator utilization and energy** (e.g., TOPS/W, TFLOPS) is fused with network metrics—packet loss, latency, throughput—into a unified “compute-network” stream feeding real-time decisions to C-PS.
Administration	* **Compute-resource provisioning:** On boot, the node advertises its compute-capability template, which OAM authenticates and syncs into the network-wide routing database. * **Service contract and billing:** multi-factor billing model pushed to edge; * **Unified orchestration:** workloads abstracted as Compute-SIDs, jointly orchestrated with network SIDs via SRv6/BGP SR Policy for cross-domain scheduling.
Maintenance	* **End-to-end quality:** ≤50 ms network (BFD/TWAMP/IOAM) + second-level compute liveness (keep-alive/health probes). * **Fault correlation and localization:** map compute-down vs. path-down to avoid false shifts; * **Self-healing:** node fail→C-PS redirects instantly.

Next Steps

- **Request more review and feedback.**
- **If necessary, consider extending OAM protocols to accommodate CATS OAM requirements.**

Thanks!