

CATS
Internet-Draft
Intended status: Standards Track
Expires: 31 July 2026

H. Fu
Q. Xiong
ZTE Corporation
B. Liu
Z. Li
China Mobile
27 January 2026

Computing-Aware Traffic Steering (CATS) Operations, Administration, and
Maintenance (OAM) Framework
draft-fu-cats-oam-fw-05

Abstract

This document describes the Operations, Administration, and Maintenance (OAM) framework and requirements for Computing-Aware Traffic Steering (CATS). The framework defines the CATS OAM layering model and functional components. It also specifies the requirements to enable fault management and performance monitoring for CATS end-to-end connections encompassing clients, network paths, and service instances.

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 31 July 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. Requirements Language	3
3. Terminology	3
4. Motivation and Problem Statement	3
5. CATS OAM Framework	5
5.1. CATS OAM Layering Model	5
5.2. CATS OAM Components	6
5.2.1. SI-OAM Component	7
5.2.2. SRV-OAM Component	8
6. CATS OAM Requirements	8
6.1. Operation	8
6.2. Administration	9
6.3. Maintenance	9
7. Security Considerations	10
8. IANA Considerations	10
9. Acknowledgements	10
10. References	10
10.1. Normative References	10
10.2. Informative References	12
Contributors	13
Authors' Addresses	13

1. Introduction

As described in [I-D.ietf-cats-usecases-requirements], edge computing provides lower response time and higher transmission rate than cloud computing by moving computing resources to the network edge. To meet the requirements of users that are highly distributive, service providers deploy the same type of service instances at multiple edge sites, which involves steering traffic from clients to the most appropriate computing instance.

Computing-Aware Traffic Steering (CATS) [I-D.ldbc-cats-framework] provides a traffic engineering approach [I-D.ietf-teas-rfc3272bis] that incorporates the dynamic states of both computing and network resources to optimize service instance selection. While such policies rely on multi-dimensional metrics to achieve service assurance, existing network-centric OAM technologies are insufficient as they focus solely on infrastructure-layer maintenance and fail to provide end-to-end visibility from the client to the service

instance. Consequently, a dedicated CATS OAM framework is required to bridge the gap between network reachability and service availability, transforming CATS from a theoretical steering logic into a manageable, carrier-grade service capable of sustaining performance in distributed computing environments.

To this end, and aligned with the architecture defined in [I-D.ldbc-cats-framework], this document specifies the OAM framework and functional requirements for CATS. It establishes a layered OAM model and defines the necessary functional components to monitor the system. Furthermore, it outlines the requirements for fault management and performance monitoring across the entire CATS service chain, covering the connection from the client through the network to the final service instance.

2. Requirements Language

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. Terminology

This document makes use of the terms defined in [I-D.ietf-cats-framework].

- * FM: Fault Management.
- * PM: Performance Monitoring.
- * SI-OAM: Service Instance OAM.
- * SVC-OAM: Service OAM.

4. Motivation and Problem Statement

Existing Operations, Administration, and Maintenance (OAM) mechanisms, such as those defined in [RFC7276], are primarily optimized for network-layer connectivity verification and path-level performance monitoring. However, the emergence of Computing-Aware Traffic Steering (CATS) [I-D.ldbc-cats-framework] introduces a multidimensional decision-making process. This necessitates an OAM framework capable of perceiving both network transport characteristics and service-level operational capabilities.

The primary objective of a CATS-specific OAM framework is to ensure that traffic steering decisions are aligned with the real-time operational status of distributed computing resources. This is critical to preventing "service black-holing," a condition where traffic is steered to a service instance that remains reachable at the network layer but is functionally unresponsive or degraded at the application level.

In the absence of a dedicated OAM framework for CATS, several critical gaps persist:

- * ***Reachability without Serviceability***: Traditional network-layer OAM protocols (e.g., Bidirectional Forwarding Detection (BFD) [RFC5880], Internet Control Message Protocol (ICMP) [RFC4443]) verify the liveness of the routing path and the node's IP interface. They lack the granularity to detect the health of the service instance hosted on that node. Consequently, a CATS Path Selector (C-PS) may continue to steer traffic to a node where the application process has crashed, deadlocked, or exceeded its resource limits, despite the node remaining network-reachable.
- * ***Temporal Staleness of Computing Metrics***: Computing metrics (e.g., CPU utilization, active thread counts, and task queue depths) exhibit significantly higher volatility compared to network topology changes. Standard control-plane advertisement or polling mechanisms often lack the necessary frequency to capture these micro-bursts, resulting in "stale" telemetry. This synchronization latency leads to sub-optimal steering decisions, potential traffic flapping, and inconsistent client experiences.
- * ***Lack of Cross-Layer Fault Demarcation***: When end-to-end (E2E) service degradation occurs (e.g., increased response latency), current OAM tools cannot natively distinguish between network-induced bottlenecks (e.g., path congestion) and compute-induced bottlenecks (e.g., resource exhaustion at the SI). This ambiguity hampers rapid fault localization, significantly increases the Mean Time to Repair (MTTR), and complicates Service Level Agreement (SLA) enforcement in multi-domain or multi-vendor environments.

- * ***Incommensurability of Heterogeneous Metrics***: Since network metrics (e.g., one-way delay, jitter, packet loss) and computing metrics (e.g., instructions per second, memory watermark) characterize distinct dimensions of service delivery, they are often collected via disjoint protocols and lack unified correlation mechanisms, such as high-precision synchronized timestamps. Without a consolidated OAM framework to harmonize these data points, the CATS Path Selector (C-PS) cannot accurately calculate the composite cost, defined as $C_{\{total\}} = f(C_{\{network\}}, C_{\{computing\}})$, thereby undermining the deterministic nature of path computation and steering efficiency.

5. CATS OAM Framework

5.1. CATS OAM Layering Model

The CATS OAM hierarchical model leverages the principles of Maintenance Domain (MD) levels, as defined in IEEE 802.1ag and ITU-T Y.1731, to extend traditional connectivity and performance management into the computing domain. This framework enables CATS-Forwarders and underlying network nodes to perform integrated anomaly detection and performance monitoring.

Based on the scope of awareness and functional granularity, the CATS OAM mechanisms are organized into four distinct layers: Link OAM, Path OAM, Instance OAM, and Service OAM. The architecture is illustrated in Figure 1.

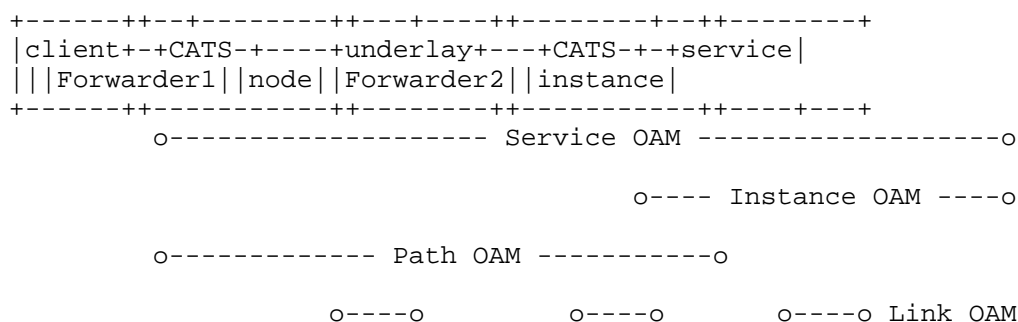


Figure 1: CATS OAM Layering Model

- * ***Link OAM***: This layer is directed at the detection of physical links or single-hop IP interfaces, representing the foundational underlay infrastructure in CATS framework [I-D.ldbc-cats-framework]. It encompasses both the internal links within the operator's infrastructure and the external interfaces interconnecting the network and computing domains. The primary

objective is to monitor the operational status between two adjacent devices to ensure fundamental IP reachability. The existing detection tools include IEEE 802.3ah, ICMP [RFC4443], and single-hop BFD [RFC5881].

- * ***Path OAM***: This layer focuses on the monitoring of transport paths between an ingress CATS-Forwarder and an egress CATS-Forwarder in CATS framework [I-D.ldb-cats-framework]. As an aggregate transport path typically carries multiple services, Path OAM is critical for ensuring that network-layer faults or resource contention from traffic multiplexing do not degrade specific service performance. Fault detection and performance monitoring are executed at the Label Switched Path (LSP) or Segment Routing (SR) path layer [RFC8402] to facilitate rapid service protection. The existing detection mechanisms include ITU-T Y.1711, MPLS Loss and Delay Measurement (LM-DM) [RFC6374], and BFD for LSP.
- * ***Instance OAM***: This layer is dedicated to the monitoring of status of computing resource and the operational health of service instances in CATS framework [I-D.ldb-cats-framework]. The metrics collected at this layer, such as CPU load and memory availability, are defined in [I-D.ietf-cats-metric-definition]. Monitoring mechanisms MUST support flexible implementation modes, including a "Push" mode, where computing nodes report dynamic load status in real-time, and a "Pull" mode, where the egress CATS-Forwarder proactively retrieves computing metrics by extending the existing OAM mechanisms.
- * ***Service OAM***: This layer provides either end-to-end or segmented visibility into the connectivity and performance between the Ingress CATS-Forwarder and the target Service Instance (SI), as defined in the CATS framework [I-D.ldb-cats-framework]. Beyond basic reachability verification, SVC-OAM is responsible for monitoring service-specific liveness, transaction success rates, and application-layer latency to ensure consistent end-to-end Quality of Experience (QoE). The supported detection mechanisms include, but are not limited to, the Two-Way Active Measurement Protocol (TWAMP) [RFC5357], application-aware probing, and HTTP-based health checks, with extensibility to accommodate emerging application-specific requirements.

5.2. CATS OAM Components

In accordance with Section 5.2 of the CATS framework [I-D.ldb-cats-framework], the CATS OAM layering model is designed to flexibly accommodate diverse OAM detection mechanisms. Consequently, this document proposes the following two new components including SI-OAM, and SVC-OAM as Figure 2 shown.

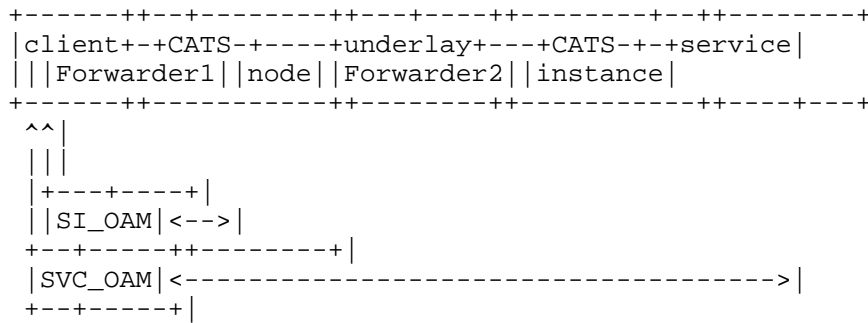


Figure 2: CATS OAM Functional Components

5.2.1. SI-OAM Component

The SI-OAM component is responsible for monitoring the operational status and computing capabilities of individual service instances. It facilitates the granular perception of service instances availability and performance. Its key functions include:

- * ***Status Monitoring***: The mechanism for periodically verifying the availability and operational state (e.g., active, inactive, or maintenance mode) of a specific service instance. This process is executed between an Egress CATS-Forwarder and its associated service instance to ensure real-time reachability and prevent traffic from being steered to an unavailable or degraded target.
- * ***Computing Metric Collection***: The process of gathering real-time computing resource telemetry from the service instance or its hosting environment. Relevant metrics include, but are not limited to, processing capacity, memory watermark, and internal queuing delay, which characterize the instantaneous capability of the instance.
- * ***Metric Reporting***: The capability to provide normalized computing telemetry to the CATS Service Metric Agent (C-SMA). This function ensures that raw resource data is translated into a consistent format to support the dynamic update of computing-aware traffic steering policies within the CATS control plane.

5.2.2. SRV-OAM Component

The SVC-OAM component provides end-to-end (E2E) visibility and performance assessment for a CATS service across the entire delivery path, originating from an Ingress CATS-Forwarder to the target service instances. Its key functions include:

- * ***Policy Verification***: Ensuring that the actual traffic forwarding path from the Ingress CATS-Router to the selected Service Instance aligns with the steering decisions made by the CATS Path Selector (C-PS).
- * ***Joint Performance Measurement***: Measuring E2E performance metrics, such as total latency (the sum of network transport time and service processing time) and jitter, to verify Service Level Agreement (SLA) compliance.
- * ***Multi-Domain Fault Isolation***: Correlating Link OAM, Path OAM, Instance OAM, and Service OAM to differentiate whether service degradation is caused by network congestion or computing resource exhaustion.

6. CATS OAM Requirements

This section specifies the OAM requirements for CATS, adhering to the operational and management guidelines defined in [I-D.draft-opsarea-rfc5706bis]. CATS OAM must bridge the gap between traditional network connectivity checks and the awareness of computing resource availability.

6.1. Operation

The operational layer is responsible for generating and collecting metrics, as well as the real-time reporting of the combined network and computing status.

- * **O-REQ-1: Multi-Dimensional Status Awareness.** The system MUST support the collection of both network metrics (latency, jitter, packet loss) and computing metrics (e.g., processing capacity, load, and availability) defined in [I-D.ietf-cats-metric-definition]. To balance precision and control plane overhead, the system SHOULD support both periodic and threshold-triggered reporting.

- * O-REQ-2: Service ID and Location Mapping. The OAM component MUST maintain the binding between a Service ID (representing the service instance) and its specific network location (Egress CATS-Forwarder). OAM probes MUST be able to target specific service instances to verify that the traffic steering policy correctly reaches the intended destination.
- * O-REQ-3: Telemetry Integration with C-SMA. All collected computing metrics SHOULD be reported to the CATS Service Metric Agent (C-SMA). This ensures that the CATS Path Selector (C-PS) has access to synchronized telemetry to perform real-time path computation and selection.

6.2. Administration

The administrative layer focuses on policy definition, security boundaries, and the configuration of steering behaviors.

- * A-REQ-1: Policy-Based Steering Configuration. The system MUST allow administrators to define CATS-specific policies, such as weighting factors for network vs. computing metrics. These policies dictate how the C-PS interprets raw telemetry when calculating the "best" service instance.
- * A-REQ-2: Differentiated Monitoring Intensity. Based on the Service Level Agreement (SLA), the system SHOULD support variable OAM intensities. High-priority services (e.g., autonomous driving or remote surgery) may require millisecond-level BFD-like monitoring, while standard web services use lower-frequency heartbeats.
- * A-REQ-3: Security and Metric Integrity. Reporting of computing and network metrics MUST be protected via encryption and authentication. This prevents malicious actors from injecting "fake" high-performance metrics to attract and intercept traffic (sinkhole attacks) or triggering oscillations in the CATS-Forwarder.

6.3. Maintenance

The maintenance layer focuses on fault isolation, performance backtracking, and ensuring the consistency of the steering state.

- * M-REQ-1: Joint Fault Demarcation. The system MUST be able to distinguish whether a service failure is caused by a network-layer issue (e.g., connectivity loss between CATS-Forwarders) or computing resource exhaustion at the Service Instance. This requires the correlation of multi-layer OAM data, including Link OAM, Path OAM, Instance OAM, and Service OAM, to accurately isolate the fault domain.
- * M-REQ-2: Historical Traceability. The OAM system SHOULD record historical snapshots of both network paths and computing status. This is critical for post-mortem analysis of "flapping" steering decisions where traffic frequently oscillates between different service instances.
- * M-REQ-3: Forwarding Plane Consistency Check. The system MUST provide mechanisms to verify that the actual traffic path taken by a packet matches the decision made by the C-PS. Any inconsistency between the intended steering policy and the actual forwarding state MUST trigger an immediate alarm and re-synchronization.

7. Security Considerations

To be discussed in future versions of this document.

8. IANA Considerations

TBD.

9. Acknowledgements

To be added upon contributions, comments and suggestions.

10. References

10.1. Normative References

[I-D.draft-opsarea-rfc5706bis]
Claise, B., Clarke, J., Farrel, A., Barguil, S.,
Pignataro, C., and R. Chen, "Guidelines for Considering
Operations and Management in IETF Specifications", Work in
Progress, Internet-Draft, draft-opsarea-rfc5706bis-06, 20
October 2025, <[https://datatracker.ietf.org/doc/html/
draft-opsarea-rfc5706bis-06](https://datatracker.ietf.org/doc/html/draft-opsarea-rfc5706bis-06)>.

[I-D.ietf-cats-metric-definition]

Yao, K., Li, C., Contreras, L. M., Ros-Giralt, J., and H. Shi, "CATS Metrics Definition", Work in Progress, Internet-Draft, draft-ietf-cats-metric-definition-04, 20 October 2025, <<https://datatracker.ietf.org/doc/html/draft-ietf-cats-metric-definition-04>>.

[I-D.ldbc-cats-framework]

Li, C., Du, Z., Boucadair, M., Contreras, L. M., and J. Drake, "A Framework for Computing-Aware Traffic Steering (CATS)", Work in Progress, Internet-Draft, draft-ldbc-cats-framework-06, 8 February 2024, <<https://datatracker.ietf.org/doc/html/draft-ldbc-cats-framework-06>>.

[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>>.

[RFC4656] Shalunov, S., Teitelbaum, B., Karp, A., Boote, J., and M. Zekauskas, "A One-way Active Measurement Protocol (OWAMP)", RFC 4656, DOI 10.17487/RFC4656, September 2006, <<https://www.rfc-editor.org/rfc/rfc4656>>.

[RFC7276] Mizrahi, T., Sprecher, N., Bellagamba, E., and Y. Weingarten, "An Overview of Operations, Administration, and Maintenance (OAM) Tools", RFC 7276, DOI 10.17487/RFC7276, June 2014, <<https://www.rfc-editor.org/rfc/rfc7276>>.

[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>>.

[RFC8402] Filsfils, C., Ed., Previdi, S., Ed., Ginsberg, L., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing Architecture", RFC 8402, DOI 10.17487/RFC8402, July 2018, <<https://www.rfc-editor.org/rfc/rfc8402>>.

[RFC8754] Filsfils, C., Ed., Dukes, D., Ed., Previdi, S., Leddy, J., Matsushima, S., and D. Voyer, "IPv6 Segment Routing Header (SRH)", RFC 8754, DOI 10.17487/RFC8754, March 2020, <<https://www.rfc-editor.org/rfc/rfc8754>>.

- [RFC9378] Brockners, F., Ed., Bhandari, S., Ed., Bernier, D., and T. Mizrahi, Ed., "In Situ Operations, Administration, and Maintenance (IOAM) Deployment", RFC 9378, DOI 10.17487/RFC9378, April 2023, <<https://www.rfc-editor.org/rfc/rfc9378>>.

10.2. Informative References

- [I-D.ietf-cats-usecases-requirements]
Yao, K., Contreras, L. M., Shi, H., Zhang, S., and Q. An, "Computing-Aware Traffic Steering (CATS) Problem Statement, Use Cases, and Requirements", Work in Progress, Internet-Draft, draft-ietf-cats-usecases-requirements-12, 12 January 2026, <<https://datatracker.ietf.org/doc/html/draft-ietf-cats-usecases-requirements-12>>.
- [I-D.ietf-teas-rfc3272bis]
Farrel, A., "Overview and Principles of Internet Traffic Engineering", Work in Progress, Internet-Draft, draft-ietf-teas-rfc3272bis-27, 12 August 2023, <<https://datatracker.ietf.org/doc/html/draft-ietf-teas-rfc3272bis-27>>.
- [RFC4443] Conta, A., Deering, S., and M. Gupta, Ed., "Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6) Specification", STD 89, RFC 4443, DOI 10.17487/RFC4443, March 2006, <<https://www.rfc-editor.org/rfc/rfc4443>>.
- [RFC5357] Hedayat, K., Krzanowski, R., Morton, A., Yum, K., and J. Babiarz, "A Two-Way Active Measurement Protocol (TWAMP)", RFC 5357, DOI 10.17487/RFC5357, October 2008, <<https://www.rfc-editor.org/rfc/rfc5357>>.
- [RFC5880] Katz, D. and D. Ward, "Bidirectional Forwarding Detection (BFD)", RFC 5880, DOI 10.17487/RFC5880, June 2010, <<https://www.rfc-editor.org/rfc/rfc5880>>.
- [RFC5881] Katz, D. and D. Ward, "Bidirectional Forwarding Detection (BFD) for IPv4 and IPv6 (Single Hop)", RFC 5881, DOI 10.17487/RFC5881, June 2010, <<https://www.rfc-editor.org/rfc/rfc5881>>.
- [RFC6374] Frost, D. and S. Bryant, "Packet Loss and Delay Measurement for MPLS Networks", RFC 6374, DOI 10.17487/RFC6374, September 2011, <<https://www.rfc-editor.org/rfc/rfc6374>>.

Contributors

Daniel Huang
ZTE Corporation
Email: huang.guangping@zte.com.cn

Cheng Huang
ZTE Corporation
Email: huang.cheng13@zte.com.cn

Wei Duan
ZTE Corporation
Email: duan.weil@zte.com.cn

Authors' Addresses

Huakai Fu
ZTE Corporation
Email: fu.huakai@zte.com.cn

Quan Xiong
ZTE Corporation
Email: xiong.quan@zte.com.cn

Bo Liu
China Mobile
Email: liubo@chinamobile.com

Zhenqiang Li
China Mobile
Email: lizhenqiang@chinamobile.com