

Computing-Aware Traffic Steering(CATS)	T. Fu, Ed.
Internet-Draft	H. Zhang, Ed.
Updates: China Academy of Information and Communications Technology	J. Wang, Ed.
2	China Mobile
0	29 May 2026
2	
6	
-	
5	
-	
2	
9	
(	
i	
f	
a	
p	
p	
r	
o	
v	
e	
d	
)	
Intended status: Informational	
Expires: 30 November 2026	

Computing-Aware Traffic Steering (CATS) Northbound Interface (NBI)  
Specification  
draft-ftzhswj-cats-northbound-api-00

Abstract

In scenarios such as industrial Internet, smart cities, and smart hospitals, enterprise customers lease a large amount of computing and network resources from operators and need to manage their own resources flexibly for agile business operations. Therefore, these customers' systems require the Computing-Aware Traffic Steering (CATS) northbound interface (NBI) to gain greater flexibility in business optimization. The CATS NBI is a standardized set of interfaces that governs interactions between the CATS system and upper-layer applications, software, and services. It defines interaction protocols, data models, message formats, and procedures. Unlike the southbound interface, which carries detailed information, the NBI simplifies data structures to deliver streamlined management capabilities. Positioned between the three-layer CATS architecture and scenario-specific applications, it is primarily invoked by user

management platforms, user business orchestration systems, AI inference platforms, and other application services. However, the current lack of a unified standard for the CATS NBI results in protocol heterogeneity, incompatible data formats, and inconsistent functionalities. This makes unified management and coordinated scheduling of multi-vendor, multi-domain CATS systems difficult, hindering large-scale deployment. A unified NBI specification is therefore urgently needed. This document defines the standard for the Computing-Aware Traffic Steering (CATS) Northbound Interface (NBI), specifying the interaction protocols, data models, message formats, and procedures between the CATS system and upper-layer management platforms and application services. As CATS technology may be adopted in future fields including the industrial Internet, Internet of Vehicles, and smart cities, enterprises and users require a concise, user-friendly, and comprehensive set of invocation interfaces. The core objective of the NBI is to enable standardized exposure of capabilities such as CATS resource query, traffic steering policy selection, metric subscription, and fault alarming, supporting unified management and coordinated scheduling of multi-vendor, multi-domain CATS systems. Based on the IETF CATS framework, metric definitions, and use case requirements documents, this document focuses on the specifications of NBI functions, protocols, and data models, providing a unified standard for CATS northbound interoperability.

#### 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 30 November 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 . . . . .	4
1.1. Background . . . . .	4
1.2. Requirements Language . . . . .	5
2. Definition of Terms . . . . .	5
3. Overall Architecture of the CATS Northbound Interface . . . . .	5
3.1. Northbound Interface Components . . . . .	6
3.2. Interaction Relationships . . . . .	6
3.3. Interface Positioning . . . . .	6
4. Interface Protocol and Transmission Specifications . . . . .	6
4.1. Basic Specifacation . . . . .	6
4.2. Transmission Security . . . . .	7
4.3. URI Naming Convention . . . . .	7
4.4. Version Compatibility . . . . .	7
5. Data Model Specifications . . . . .	7
5.1. Data Modeling Language . . . . .	7
5.2. Core Data Models . . . . .	7
6. Northbound Interface Functions . . . . .	7
6.1. Resource Management Module . . . . .	8
6.2. Policy Management Module . . . . .	8
6.3. Metric Subscription Module . . . . .	8
6.4. Alarm Management Module . . . . .	8
6.5. Configuration Management Module . . . . .	8
7. Message Format and Interaction Procedures . . . . .	8
7.1. General Message Format . . . . .	8
7.2. 8.2 Typical Interaction Procedures . . . . .	9
8. Interface Security Requirements . . . . .	9
9. IANA Considerations . . . . .	9
10. Security Considerations . . . . .	9
11. References . . . . .	9
11.1. Normative References . . . . .	9
11.2. Informative References . . . . .	9
Appendix A. Appendix 1 . . . . .	9
Acknowledgements . . . . .	10
Contributors . . . . .	10
Authors' Addresses . . . . .	10

## 1. Introduction

### 1.1. Background

With the rapid proliferation of low-latency, high-computing services such as agent collaboration, edge computing, AR/VR, and intelligent transportation, traditional traffic steering mechanisms that rely solely on network connectivity fail to adapt to dynamic changes in computing resources. This easily leads to issues including computing overload, unbalanced scheduling, and degraded user experience. As a new traffic engineering technology, Computing-Aware Traffic Steering (CATS) fundamentally integrates multi-dimensional metrics such as computing status, network quality, and service load to dynamically select optimal service instances and steer traffic, ensuring service Quality of Experience (QoE). The CATS system consists of core components including the C-SMA (CATS Service Metric Agent), C-NMA (CATS Network Metric Agent), C-PS (CATS Path Selector), and C-Forwarder. It requires exposing a unified interface to upper-layer management platforms and application services to enable resource control, policy configuration, metric visualization, and fault operation and maintenance. Currently, there is a lack of a unified standard for the CATS northbound interface, resulting in protocol heterogeneity, incompatible data formats, and inconsistent functionalities. This may prevent users from gaining sufficient capabilities when managing their CATS systems across vendors, hindering large-scale deployment. In scenarios such as the industrial Internet, smart cities, and smart hospitals, enterprise customers lease substantial computing and network resources from operators and require flexible resource management to support agile business operations. Therefore, these customers' systems need the Computing-Aware Traffic Steering (CATS) northbound interface to achieve greater flexibility in business optimization. The CATS northbound interface is a standardized set of interfaces that governs interactions between the CATS system and upper-layer applications, software, and services. It defines interaction protocols, data models, message formats, and procedures. Unlike the southbound interface, which carries detailed information, it delivers streamlined management capabilities by simplifying interface data structures. Positioned between the three-layer CATS architecture and scenario-specific applications, it is primarily invoked by user management platforms, user business orchestration systems, AI inference platforms, and other application services. However, the current lack of a unified standard interface for CATS leads to protocol heterogeneity, incompatible data formats, and inconsistent functionalities. This makes unified management and coordinated scheduling of multi-vendor, multi-domain CATS systems difficult, restricting large-scale deployment. A unified northbound interface specification is therefore urgently needed.

## 1.2. 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.

## 2. Definition of Terms

Computing-Aware Traffic Steering(CATS): A traffic engineering technique that dynamically steers traffic by jointly considering computing status, network quality, and service load. Computing-Aware Traffic Steering Metric(CATS Metric):A multi-dimensional indicator used to evaluate computing resources, network performance, and service status for CATS-based traffic decisions. CATS Service Metric Agent(C-SMA):A CATS component responsible for collecting, aggregating, and reporting service-related metrics such as load and QoE. CATS Network Metric Agent(C-NMA):A CATS component responsible for collecting, aggregating, and reporting network-related metrics such as latency, packet loss, and bandwidth. CATS Path Selector(C-PS):A CATS core component that selects optimal service instances and forwarding paths based on aggregated CATS metrics. CATS Service ID(CS-ID):A unique identifier assigned to a CATS service for service-level management and policy binding. CATS Service Contact Instance ID(CSCI-ID):A unique identifier for a specific instance of a CATS service, used for instance-level scheduling and monitoring. CATS Northbound Interface(CATS NBI):A standardized interface set enabling interaction between the CATS system and upper-layer management platforms and applications. Operations, Administration, and Maintenance(OAM):A set of functions for network/system monitoring, fault management, performance measurement, and maintenance. Quality of Experience(QoE):A holistic metric reflecting the end-user's perceived quality of a service, considering latency, jitter, packet loss, and service availability. REST Configuration(RESTCONF):An HTTP/JSON-based protocol defined by IETF RFC 8040 for configuring and retrieving data modeled in YANG. YANG: A data modeling language defined by IETF RFC 7950, used to model configuration and operational data for network and service management.

## 3. Overall Architecture of the CATS Northbound Interface

### 3.1. Northbound Interface Components

The CATS northbound interface consists of three components, listed from top to bottom as follows: Application Adaptation Module: Connected to upper-layer management platforms, business orchestration systems, and application services (e.g., AR/VR, AI inference platforms), acting as the northbound interface callers. Interface Service Module: The core functions of the CATS northbound interface, including protocol adaptation, data model parsing, functional logic processing, and security control modules, serving as the core of this standard. CATS Adaptation Module: Connected to internal components of the CATS system, which receive northbound interface instructions and execute scheduling logic. To simplify user's operations, it's mainly used by C-SMA.

### 3.2. Interaction Relationships

The interaction modes of the northbound interface are divided into synchronous request-response and asynchronous subscription-push: Synchronous interaction: The upper layer initiates requests for query, configuration, and policy delivery, and the CATS interface returns responses synchronously (e.g., resource query, policy configuration). Asynchronous interaction: The upper layer subscribes to metrics and alarms, and the CATS interface pushes data in real time (e.g., computing load, fault alarms).

### 3.3. Interface Positioning

Unique Entry Point for NBI: All upper-layer CATS-related operations access the system through the northbound interface, enabling unified management and control. Stateless Interaction: Interface interactions are session-independent, supporting high availability and load balancing. Cross-Domain Interoperability: It supports unified northbound access for both single-domain and multi-domain CATS systems, adapting to diverse scenarios of carriers and enterprises.

## 4. Interface Protocol and Transmission Specifications

### 4.1. Basic Specifacation

The CATS northbound interface mandatorily adopts the RESTCONF protocol (IETF RFC 8040), transmitted over HTTP with a unified JSON data format and compatibility with YANG data models.

#### 4.2. Transmission Security

Interface transmission mandatorily uses TLS 1.3 encryption; plaintext HTTP transmission is prohibited. Port: Default port 443, with support for custom ports. Certificate: X.509 certificates are used for mutual authentication to ensure the authenticity of both communicating parties

#### 4.3. URI Naming Convention

The northbound interface URI follows a hierarchical naming rule.

#### 4.4. Version Compatibility

Interface versions adopt path-based versioning (v1/v2) to support backward compatibility: Legacy interface versions remain available, while new features in later versions use independent paths. When extending data models, newly added fields are optional by default and do not affect parsing of older versions.

### 5. Data Model Specifications

#### 5.1. Data Modeling Language

The core data models are defined using YANG 1.1 (IETF RFC 7950), supporting JSON format mapping. They reuse the IETF CATS metric model (draft-ietf-cats-metric-definition) and add northbound-specific fields.

#### 5.2. Core Data Models

- 6.2.1 CATS Resource Model (cats-resource.yang) TBA.
- 6.2.2 CATS Policy Model (cats-policy.yang) TBA.
- 6.2.3 CATS Metric Model Adopts the three-level CATS metric model defined in draft-ietf-cats-metric-definition.
- 6.2.4 CATS Alarm Model (cats-alarm.yang) TBA.

### 6. Northbound Interface Functions

### 6.1. Resource Management Module

Provides query capabilities for CATS sites, service instances, and computing/network resources:

- a. Shall support querying the list and status of all service sites within user permissions.
- b. Shall support querying computing resources (CPU/GPU/utilization, computing score) and network resources (latency/packet loss/bandwidth, network score) of a specified site.
- c. Shall support querying the status, load, and QoE score of a service instance identified by a specified CS-ID/CSCI-ID.
- d. May support querying global Level 2 aggregated metrics of CATS.
- e. TBD.

### 6.2. Policy Management Module

Supports querying, selecting, modifying, deleting, enabling/disabling traffic steering policies for each CATS service:

- a. Shall support creating static/dynamic/hybrid steering policies (binding CS-ID, priority, trigger threshold, target instance).
- b. Shall support querying all policies or details of a specified policy.
- c. Shall support updating policy trigger conditions, priority, and target instances.
- d. Shall support enabling/disabling a specified policy.
- e. Shall support deleting invalid policies.
- f. TBD.

### 6.3. Metric Subscription Module

TBD.

### 6.4. Alarm Management Module

TBD.

### 6.5. Configuration Management Module

TBA.

## 7. Message Format and Interaction Procedures

### 7.1. General Message Format

#### 8.1.1 Success Response (JSON) TBA.



## 7.2. 8.2 Typical Interaction Procedures

8.2.1 Resource Query (Synchronous GET) TBA.

8.2.2 Policy Provisioning (Synchronous POST) TBA.

8.2.3 Metric Subscription (Asynchronous POST) TBA.

8.2.4 Alarm Subscription (Asynchronous POST) The procedure is the same as metric subscription. Alarm data is pushed in real time upon successful subscription.

## 8. Interface Security Requirements

TBA.

## 9. IANA Considerations

This memo includes no request to IANA.

## 10. Security Considerations

This document should not affect the security of the Internet.

## 11. References

### 11.1. Normative References

[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/info/rfc8174>>.

### 11.2. Informative 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/info/rfc2119>>.

[exampleRefMin]  
Surname, Initials., "Title", 2006.

[exampleRefOrg]  
Organization, "Title", 1984, <<http://www.example.com/>>.

## Appendix A. Appendix 1

TBD.

## Acknowledgements

TBD.

## Contributors

Thanks to all of the contributors.

## Authors' Addresses

Fu Tao (editor)  
China Academy of Information and Communications Technology  
Huayuanbei No.52  
beijing  
beijing, 100191  
China  
Email: futao@caict.ac.cn

Zhang Hengsheng (editor)  
China Academy of Information and Communications Technology  
Huayuanbei No.52  
beijing  
beijing, 100191  
China  
Email: zhanghengsheng@caict.ac.cn

Wang Jing (editor)  
China Mobile  
beijing  
beijing, 100191  
China  
Email: wangjingjc@chinamobile.com