

CATS  
Internet-Draft  
Intended status: Standards Track  
Expires: 16 April 2026

C. Li  
Huawei Technologies  
Z. Du  
China Mobile  
J. Drake  
Independent  
Y. Shang  
Tsinghua University  
13 October 2025

Computing-Aware Traffic Steering (CATS) Using Segment Routing  
draft-lbdd-cats-dp-sr-06

Abstract

This document describes a solution that adheres to the Computing-Aware Traffic Steering (CATS) framework. The solution uses anycast IP addresses as the CATS service identifiers and Segment Routing (SR) as the data plane encapsulation to achieve computing-aware traffic steering among multiple services 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 16 April 2026.

Copyright Notice

Copyright (c) 2025 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. Terminology . . . . .	3
3. Solution Overview . . . . .	3
3.1. Realization of CATS Framework Components . . . . .	3
3.1.1. CATS Identifiers . . . . .	3
3.1.2. CATS Components . . . . .	3
3.2. Realization of the CATS Framework Workflow . . . . .	4
3.2.1. Service Announcement . . . . .	4
3.2.2. Metrics Distribution . . . . .	4
3.2.3. Service Access Processing . . . . .	4
3.2.4. Service Instance Affinity . . . . .	7
3.3. Operational and Manageability Considerations . . . . .	7
4. Security Considerations . . . . .	8
5. IANA Considerations . . . . .	8
6. Acknowledgements . . . . .	8
7. References . . . . .	8
7.1. Normative References . . . . .	9
7.2. Informative References . . . . .	9
Contributors . . . . .	10
Authors' Addresses . . . . .	10

## 1. Introduction

As described in [I-D.ietf-cats-usecases-requirements], traffic steering that takes into account computing resource metrics would benefit several services, e.g., the latency-sensitive service. A typical example would be the immersive services that rely upon the use of augmented reality or virtual reality (AR/VR) techniques.

[I-D.ietf-cats-framework] defines a framework for Computing-Aware Traffic Steering (CATS). Such a framework defines an approach for making compute- and network-aware traffic steering decisions in networking environments where services are deployed in many locations.

The CATS framework is an overlay framework for the selection of the suitable service contact instance for placing a service request. The exact characterization of 'suitable' will be determined by a combination of networking and computing metrics. The CATS framework does not assume any specific data plane and control plane solutions.

This document proposes a data plane solution for the realization of CATS. The solution uses an anycast IP address as the Computing-aware Service ID (CS-ID) associated with a service. Also, the solution uses Segment Routing (SR) as the data plane encapsulation from an Ingress CATS-Router to an Egress CATS-Router.

## 2. Terminology

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

Note: Terms such as CATS Service Contact Instance ID (CSCI-ID) have been updated in the CATS framework [I-D.ietf-cats-framework].

## 3. Solution Overview

This section describes the details of realizing CATS identifiers, CATS components, and realted workflow.

### 3.1. Realization of CATS Framework Components

#### 3.1.1. CATS Identifiers

A CATS Service ID (CS-ID) is an anycast IPv4 or IPv6 address. Such an IP address is associated with a specific service that is reachable via one or multiple service contact instances.

The CATS overlay encapsulation is established from an Ingress CATS-Router to an Egress CATS-Router connected to a service contact instance. The service contact instance is typically hosted in a service site.

As defined in the CATS framework, a CSCI-ID is the identifier of a specific service contact instance. Depending on the deployment requirements, a CSCI-ID may be needed to indicate where to forward the packet in the case that multiple sites connect to the same Egress CATS-Router.

#### 3.1.2. CATS Components

In the context of this document, CATS-Routers are required to support SR encapsulation, including SR-MPLS [RFC8660] and SRv6 [RFC8986].

The CATS Traffic Classifier (C-TC) is assumed to be running on Ingress CATS-Routers.

For each service site, one or multiple C-SMAs can be implemented within the site to collect the metrics of the service instances.

### 3.2. Realization of the CATS Framework Workflow

#### 3.2.1. Service Announcement

The service's anycast IP address may be announced by using a rendezvous service (DNS, for example). Clients can obtain the CS-ID of the service from the rendezvous service used by the application. It is out of scope of this document to provide a comprehensive list of all candidate rendezvous services.

#### 3.2.2. Metrics Distribution

As per the CATS framework, CS-ID routes with metrics are distributed among the overlay CATS Routers. The detailed control plane solutions of metrics distribution are out of the scope of this document. However, a sample procedure is provided for the readers' convenience.

For example, BGP can be used to distribute CS-ID routes with metrics.

In the case of the C-SMA running as a stand-alone entity outside an Egress CATS-Router, the C-SMA collects the metrics of computing resource within a service site and distributes the CS-ID routes with the collected metrics to the Egress CATS-Router. Egress CATS-Routers will generate the new metrics combined with network metrics and computing-related metrics, and redistribute the CS-ID routes to Ingress CATS-Routers. In the case of the C-SMA running as a logic entity on an Egress CATS-Router, the same process will be performed inside the Egress CATS-Router.

As described in Section 3.4 of [I-D.ietf-cats-framework], CATS can be deployed in a distributed model, a centralized model, or a hybrid model. In a centralized model or a hybrid model, the routes with metrics may be collected by centralized controllers. BGP-LS may be a candidate solution to collect the route with metrics from CATS-Routers to controllers; the use of BGP-LS is however out of the scope of this document.

A centralized controller may also install forwarding policies on Ingress CATS-Routers to steer the traffic; how these policies are communicated to the routers is out of the scope of this document.

#### 3.2.3. Service Access Processing

Two SR [RFC8402] data plane approaches are supported: SRv6 [RFC8986] and SR-MPLS [RFC8660]. This section introduces a solution based upon SRv6 and SR-MPLS as data planes for CATS purposes.

An Ingress CATS-Router generates SRv6/SR-MPLS encapsulations from itself to Egress CATS-Routers according to the SR policy received from a controller. An Ingress CATS-Router receives service routes with network and computing-related metrics from Egress CATS-Routers. An C-PS will select the best service site according to the received service routes and routing policies. Once the best service site is selected, the associated Egress CATS-Router can be determined and the appropriate SR encapsulation from an Ingress CATS-Router to the C-PS-computed Egress CATS-Router can be selected.

When a service access packet is received by an Ingress CATS-Router, it is classified by the C-TC component. When a matching classification entry is found for this service access packet, the Ingress CATS-Router encapsulates and forwards it to the C-PS selected Egress CATS-Router via the matching SR tunnel.

#### 3.2.3.1. SRv6

As shown in Figure 1, SRv6 tunnels are established from Ingress CATS-Routers to Egress CATS-Routers.

There may be multiple encapsulations from a single Ingress CATS-Router to different Egress CATS-Routers so that the ingress can choose the best Egress CATS-Router connected to the target site.

Furthermore, there may be multiple tunnels from a single Ingress CATS-Router to a single Egress CATS-Router, e.g., to provide different connectivity performance guarantees.

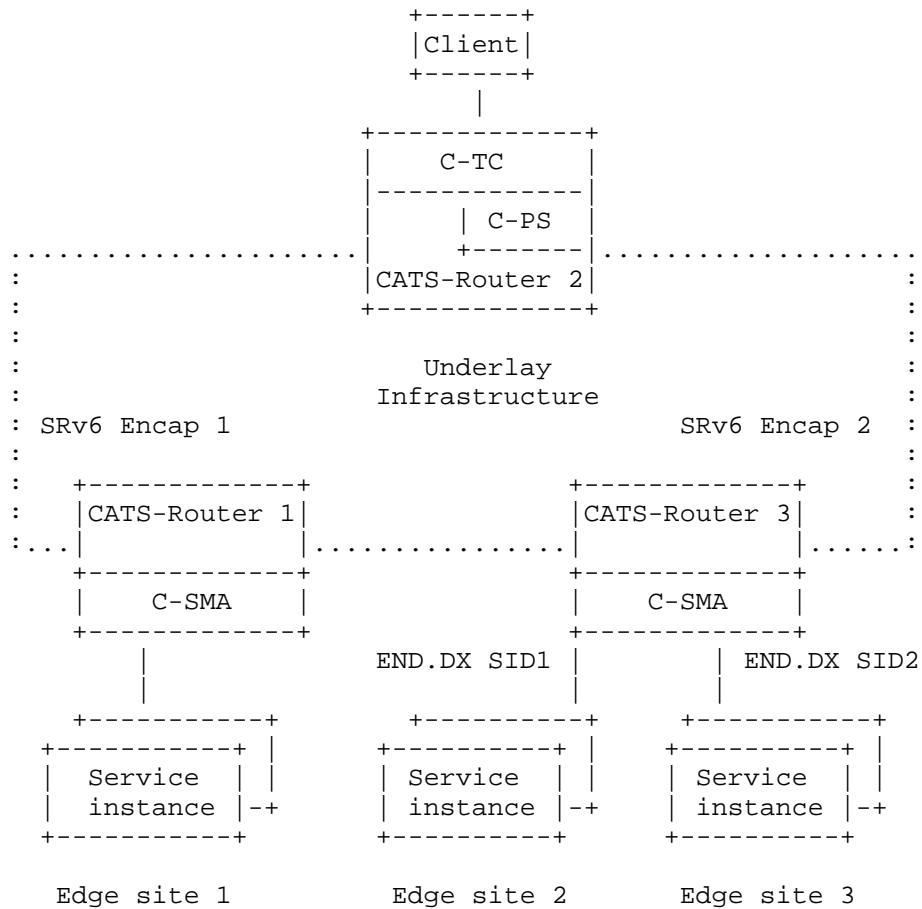


Figure 1: Using SRv6 in CATS

In some cases, multiple service sites may be connected to a single Egress CATS-Router. To demux these sites, a specific attachment circuit must be provided to indicate the specific target service. In order to explicitly indicate the interface towards a site, an END.DX [RFC8986] is encoded as the last segment in the SRv6 encapsulation. The associated END.DX is learned from the control plane.

When the traffic reaches the Egress CATS-Router, the SRv6 packet is decapsulated and the traffic is forwarded to the service contact instance. How the packet is handled beyond that point is out of the scope.

#### 3.2.3.2. SR-MPLS

Similarly, SR-MPLS can be used as the overlay CATS encapsulation. The forwarding path is encoded as an MPLS label stack, and a potential VPN label can be included as the last label to indicate to steer the traffic through a specific interface to a target service contact instance in the case multiple service sites connect to the same Egress CATS-Router.

#### 3.2.4. Service Instance Affinity

As per [I-D.ietf-cats-framework], different services may have different notions of what constitutes a 'flow' and may thus identify a flow differently. Typically, a flow is identified by the 5-tuple transport coordinates (source and destination addresses, source and destination port numbers, and protocol).

For a service that requires service instance affinity, the Ingress CATS-Router needs to forward all the packets in a flow to the same service instance. Section 3.2.3 describes the general procedure of how to steer the packets of a flow to the same SR tunnel. When the packet is the first packet in the flow, the flow characteristics might not be matched in the C-TC, and a forwarding entry will be created for this flow. When the flow characteristics can be matched in the C-TC, the packet will be forwarded to the same tunnel selected by the previous packet in the flow, so that the service instance affinity can be guaranteed.

### 3.3. Operational and Manageability Considerations

For the routes of the anycast IP address, there may be multiple candidate routes on the Ingress Router, which have different Egress routers as the next hop. According to related computing metrics and network metrics, each candidate route can be associated with a dynamic weight.

There may also be multiple SR policies between the Ingress router and the target Egress router. For a CATS service, it should have an intent for the selecting or mapping of an SR policy. For example, the intent or objective can be low-latency, which appears as the color in an SR policy tuple <Headend, Color, Endpoint> [RFC9256].

After a service contact instance or an Egress Router is selected for a CATS flow considering weights of candidate routes, the Ingress CATS-Router needs to associate the flow with a proper SR policy between the Ingress CATS-Router and the Egress CATS-Router.

Some accounting requirements are listed below to record the amount of CATS traffic in the operation.

- \* An Ingress router MUST be able to account the amount of the CATS traffic along the selected SR policy.
- \* An Egress Router MUST be able to account the amount of the CATS traffic received from a selected SR policy.

The weight of a route will be changed in operation, therefore, some weight changing requirements are listed below. It is assuming that multiple service instances in different service sites form a load balance group at the Ingress router for a CATS service.

- \* Large weight SHOULD be configured to the route to the service site that can serve more clients.
- \* When the service site is busy, for example, certain essential recourse is about to be exhausted, the related route for it on the Ingress Router should lower its weight, or even leave the load balance group temperately.
- \* If the latency of a specific SR tunnel bearing the CATS traffic exceeds a threshold, its related route on the Ingress Router should lower its weight, or even leave the load balance group temperately.

#### 4. Security Considerations

This document specifies a CATS solution using anycast IP addresses as CS-IDs and SR as data plane. It does not introduce further security threats considering to the existing ones in [RFC8402], [RFC8660], [RFC8986] and [I-D.ietf-cats-framework].

Anycast-related security considerations are discussed in Section 4.4 of [RFC7094].

#### 5. IANA Considerations

This document makes no requests for IANA action.

#### 6. Acknowledgements

Many thanks to Mohamed Boucadair for his valuable help on this document.

#### 7. References



## 7.1. Normative References

- [I-D.ietf-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-ietf-cats-framework-15, 15 September 2025, <<https://datatracker.ietf.org/doc/html/draft-ietf-cats-framework-15>>.
- [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/info/rfc8402>>.
- [RFC8660] Bashandy, A., Ed., Filsfils, C., Ed., Previdi, S., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing with the MPLS Data Plane", RFC 8660, DOI 10.17487/RFC8660, December 2019, <<https://www.rfc-editor.org/info/rfc8660>>.
- [RFC8986] Filsfils, C., Ed., Camarillo, P., Ed., Leddy, J., Voyer, D., Matsushima, S., and Z. Li, "Segment Routing over IPv6 (SRv6) Network Programming", RFC 8986, DOI 10.17487/RFC8986, February 2021, <<https://www.rfc-editor.org/info/rfc8986>>.
- [RFC9256] Filsfils, C., Talaulikar, K., Ed., Voyer, D., Bogdanov, A., and P. Mattes, "Segment Routing Policy Architecture", RFC 9256, DOI 10.17487/RFC9256, July 2022, <<https://www.rfc-editor.org/info/rfc9256>>.

## 7.2. 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-08, 13 October 2025, <<https://datatracker.ietf.org/doc/html/draft-ietf-cats-usecases-requirements-08>>.
- [RFC7094] McPherson, D., Oran, D., Thaler, D., and E. Osterweil, "Architectural Considerations of IP Anycast", RFC 7094, DOI 10.17487/RFC7094, January 2014, <<https://www.rfc-editor.org/info/rfc7094>>.

## Contributors

Dirk Trossen  
Huawei Technologies  
Email: dirk.trossen@huawei.com

Luigi Iannone  
Huawei Technologies  
Email: luigi.iannone@huawei.com

Yizhou Li  
Huawei Technologies  
Email: liyizhou@huawei.com

Hang Shi  
Huawei Technologies  
Email: shihang9@huawei.com

## Authors' Addresses

Cheng Li  
Huawei Technologies  
China  
Email: c.l@huawei.com

Zongpeng Du  
China Mobile  
China  
Email: duzongpeng@chinamobile.com

John E Drake  
Independent  
United States of America  
Email: je\_drake@yahoo.com

Yuxiang Shang  
Tsinghua University  
China  
Email: shang-yx24@mails.tsinghua.edu.cn