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

X. Yi, Ed.
N. Zhang, Ed.
China Unicom

12 September 2025

Hierarchical methods of computing metrics distribution
draft-yi-cats-hierarchical-metric-distribution-02

Abstract

This document analyzes the necessarity of setting hierarchical methods of computing metrics distribution. Besides, we propose the workflow of hierarchical metric distribution for different CATS frameworks.

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 March 2026.

Table of Contents

1. Introduction

The path computation needs to process both network and computing metrics in CATS. As described in [I-D.ietf-cats-framework], after the computing metrics are collected, they will be distributed to CATS Path Selector (C-PS) for path computation. The C-PS can be integrated into ingress CATS-Forwarders or deployed as a functional component of a centralized controller. Computing metrics are defined in [I-D.ietf-cats-metric-definition]. And [I-D.ietf-cats-usecases-requirements] describes the importance of the representation and encoding of computing metrics. As service scale increases, the data amount reported to C-PS and size of routing table increases, which will bring greater processing pressure to network devices. As a result, the service quality will be affected. Therefore, it is necessary to set a hierarchical method of computing metric distribution between service site and different C-PSes. The hierarchical method can be set according to geographical distance or performance of service sites, etc.

1.1. Terminology

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. Conventions and Definitions

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. This document uses terms defined in [I-D.ietf-cats-framework]. We list them below for clarification.

*Computing-Aware Traffic Steering (CATS): An architecture that takes into account the dynamic nature of computing resources and network state to steer service traffic to a service instance. This dynamicity is expressed by means of relevant metrics.

*Ingress CATS-Forwarder: An entity that steers service-specific traffic along a CATS-computed path that leads to an Egress CATS-Forwarder that connects to the most suitable service site that host the service contact instance selected to satisfy the initial service request.

*Egress CATS-Forwarder: An entity that is located at the end of a CATS-computed path and which connects to a CATS-serviced site.

*CATS Path Selector (C-PS): A functional entity that selects paths towards service locations and instances and which accommodates the requirements of service requests. Such a path selection engine takes into account the service and network status information.

3. Definition and requirements of hierarchical metric distribution

The hierarchical method of cpmputing metrics is based on the CATS framework defined in [I-D.ietf-cats-framework](see Figure 1, the figure is replicated here for better understanding).

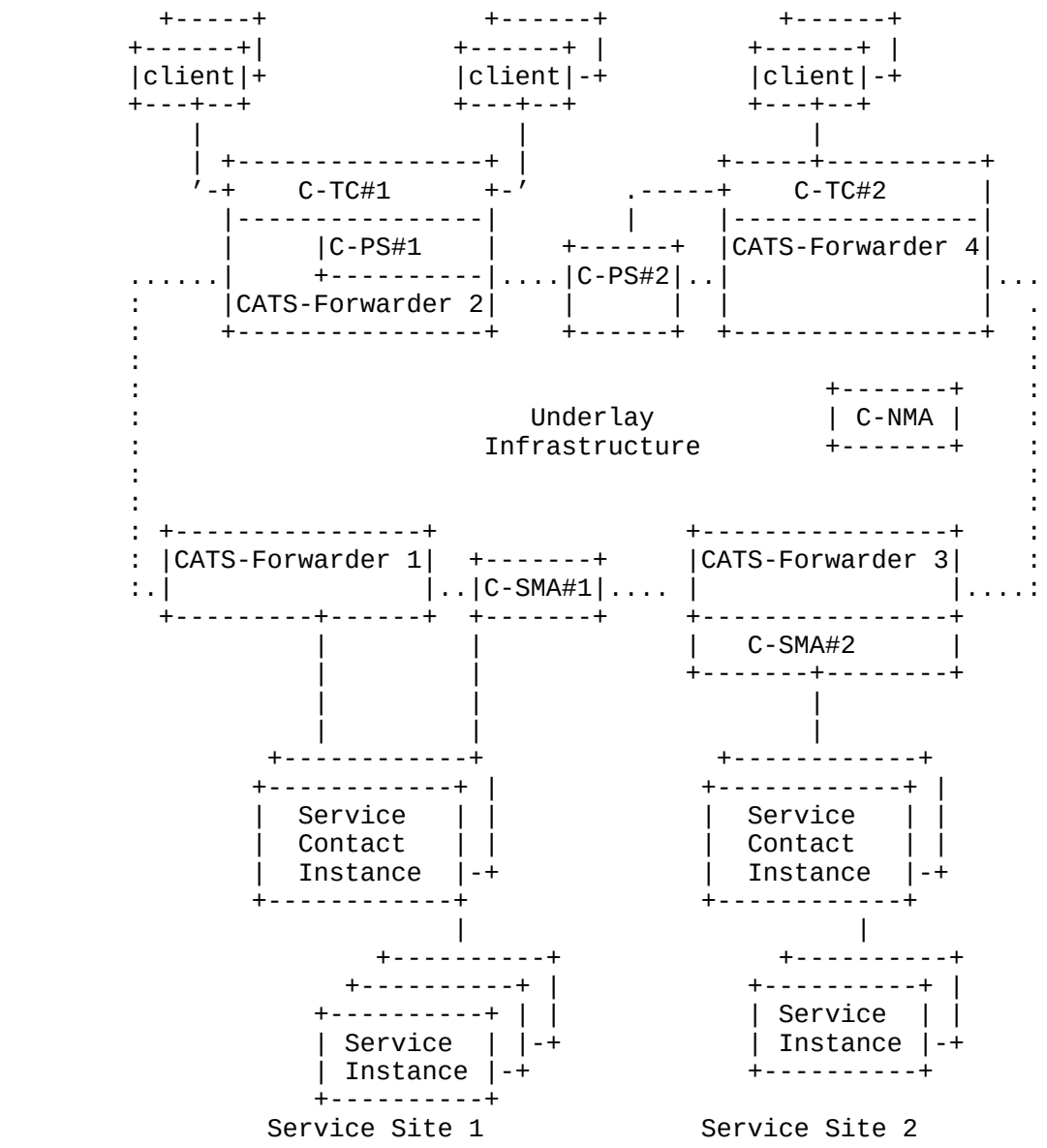

Figure 1: CATS-Functional-Components

The hierarchical metric distribution is defined as reporting different levels of the computing metric to C-PS in ingress CATS-Forwarder according to geographical scope (distance or radius) and performance of service sites, etc. The service sites classified as high priority report high-level computing metrics (full metric

information). And the service sites classified as low priority only report a simple computing metric (i.e. available or unavailable). If the priorities are classified by distance, then the near service site can be set to a higher priority. On the one hand, setting a hierarchical method of computing metric distribution can reduce the processing pressure of network devices and the size of distribution data and routing table. On the other hand, the increase in service delay caused by processing large amounts of data can be avoided. The impact is particularly obvious when C-PS is deployed in CATS-Forwader. Therefore, hierarchical method of computing metric distribution is particularly important in distributed and hybrid model.

In the distributed model, all computing metrics are processed in C-PS in the ingress CATS-Forwarder. It is better to set a suitable priority for each service site. Therefore, the processing pressure of the CATS-Forwarder can be relieved while maintaining sufficient service resource utilization.

In the centralized model, all computing metrics are processed in centralized C-PS. A centralized C-PS may have a hierarchical control system, and different levels of controllers can be set to receive different levels of the computing metric.

In the hybrid model, computing metrics can be chosen to be processed in the ingress CATS-Forwarder or centralized C-PS according to service types. [I-D.yi-cats-hybrid-solution] has discussed how to choose the appropriate processing location in different scenarios. Because not all computing metrics need to be processed in the CATS-Forwarder in hybrid model, the hierarchical method can be more flexible.

4. Optional hierarchical parameters

The priority of service sites is determined by parameters of service sites, such as geographical distance and performance. When calculating the priority, different parameters of service sites are assigned different weights based on the service type. The parameter scores of different service site parameters are obtained first. The sum of all products of parameter scores and corresponding weights is the priority score.The higher the priority score, the higher the priority of the service site.

4.1. Geographical distance

Geographical distance is the distance between ingress CATS-Forwarder and egress CATS-Forwarder. If prioritizing based on geographic distance, whether the distance between each network device is known needs to be considered. If the distance is unknown, the distance can be measured through the network packet transmission delay. Parameter score of distance is negatively correlated with the distance.

4.2. Performance of service sites

The performance of a service site is related to the processing speed and resource idle rate of the service site. Parameter score of service site performance is positively correlated with the performance.

4.3. Number of service sites

TBD

4.4. Other Options

TBD

5. Workflow of hierarchical metric distribution

The ingress CATS-Forwarder receives service request and send it to egress CATS-Forwarder connected to the service site. Then the egress CATS-Forwarder calculates the priority of service sites according to the obtained parameters of service sites and send instruction messages to them. Instruction messages are used to indicate the priority of the corresponding service site. The egress CATS-Forwader recieves computing metrics from service sites and send computing metrics to C-PS (CATS Path Selector). The computing metrics of a service site is determined based on the priority of the service site.. For different CATS frameworks, the workflow of hierarchical metric distribution is different.

5.1. Distributed model

When an egress CATS-Forwarder establishes a one-to-one neighbor relationship with an ingress CATS-Forwarder, the egress CATS-Forwader determines what level of computing metric to report to C-PS in this ingress CATS-Forwarder.

5.2. Centralized model

The egress CATS-Forwader determines what level of computing metric to report to the different level of controller in the centralized C-PS.

5.3. Hybrid model

The hierarchical policies are set in both egress CATS-Forwarder and centralized C-PS.

Firstly, the egress CATS-Forwarder determine whether to report the collected computing metrics to centralized C-PS or C-PS in the ingress CATS-Forwarder according to the requirements (delay or global utilization, etc.) of different service types.

If the CATS-Forwarder is selected, the hierarchical metric distribution workflow follows the distributed model mentioned above.

If there is hierarchical control in hybrid model, the hierarchical metric distribution workflow follows the centralized model mentioned above.

6. Security Considerations

TBD

7. IANA Considerations

TBD

8. Normative 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>.

[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>.

[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-13, 3 September 2025, <https://datatracker.ietf.org/doc/html/draft-ietf-cats-framework-13>.

[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-07, 10 June 2025, <https://datatracker.ietf.org/doc/html/draft-ietf-cats-usecases-requirements-07>.

[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-03, 7 July 2025, <https://datatracker.ietf.org/doc/html/draft-ietf-cats-metric-definition-03>.

[I-D.yi-cats-hybrid-solution]
Yi, X., Pang, R., and H. Shi, "Hybrid Computing and Network Awareness and Routing Solution for CATS", Work in Progress, Internet-Draft, draft-yi-cats-hybrid-solution-03, 24 July 2024, <https://datatracker.ietf.org/doc/html/draft-yi-cats-hybrid-solution-03>.

Authors' Addresses

Xinxin Yi (editor)
China Unicom
Beijing
China
Email: yixx3@chinaunicom.cn

Naihan Zhang (editor)
China Unicom
Beijing
China
Email: zhangnh12@chinaunicom.cn