

CATS
Internet-Draft
Intended status: Standards Track
Expires: 14 May 2026

Z. Li
J. Mao
Huawei
10 November 2025

Application Aware Computing Network
draft-li-cats-application-aware-computing-network-02

Abstract

This document describes a solution framework that adheres to the CATS framework. The solution uses APN as part of the CATS service identifier and flow identifier.

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 14 May 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
2.1. Requirements Language	3
3. Locator and APN ID	3
3.1. APN Segment	3
3.2. LOC + APN ID	4
4. Overview	4
4.1. Realization of CATS Framework Components	4
4.2. Realization of CATS Framework Workflow	5
4.2.1. Service Announcement	5
4.2.2. Metric Distribution	5
4.2.3. Service Request Processing	5
4.2.4. Service Instance Affinity	5
5. APN for Real Locator	6
6. Security Considerations	6
7. IANA Considerations	6
8. Acknowledgements	6
9. References	6
9.1. Normative References	6
9.2. Informative References	7
Authors' Addresses	7

1. Introduction

Many services deploy their service instances in multiple geographically distributed sites to get better response time. As described in [I-D.ietf-cats-usecases-requirements], traffic steering that takes into account both the computing resource metric and network metric would improve the QoE of several services, e.g., AR/VR and intelligent transportation.

A CATS framework is described in [I-D.draft-ldbc-cats-framework]. It defines following core concepts:

1. CATS service identifier represents a service which consists of multiple service contact instances.
2. Service contact instance affinity means that packet that belongs to a flow should always goes to the same service contact instance through the same path.

These concepts are similar to APN described in [I-D.draft-li-apn-header]. The Application-aware Networking (APN) framework [I-D.li-apn-framework] defines that application-aware information (i.e. APN attribute) including APN identification (ID) and/or APN parameters (e.g. network performance requirements) is

encapsulated at network edge devices and carried in packets traversing an APN domain in order to facilitate service provisioning, perform fine-granularity traffic steering and network resource adjustment. [I-D.li-apn-app-side-framework] defines the extension of the APN framework for the application side. In this extension, the APN resources of an APN domain is allocated to applications which compose and encapsulate the APN attribute in packets. The APN ID includes application group ID and user group ID. Application group ID can be used as part of CATS service identifier. User group ID + application group ID can be used as CATS flow ID. This document describes a CATS framework using APN. The realization can use APN framework for application side.

2. Terminology

This document reuses terms defined in [I-D.draft-ldbc-cats-framework] and [I-D.draft-li-apn-header].

2.1. 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. Locator and APN ID

In CATS, the packet needs to carry the locator information and the application identifier information. The application identifier information can be carried in APN ID. As described in [I-D.draft-li-apn-header], the length of the APN ID can be 32bit, 64 bit or 128bit. Depending on whether APN ID can fit into 128bit, the following two formats are defined. They carry the same information and can be used in different deployment scenarios.

3.1. APN Segment

APN segment is an SRv6 SID[RFC8986], consisting of LOC:FUNCT:ARG, where:

- * LOC is routable and leads to the service contact instance.
- * FUNCT is APN ID which consists of user group id and application group ID. The length of APN ID is 32 bit or 64 bit.

3.2. LOC + APN ID

If 128 bit is not enough to hold the APN ID + LOC. Then the APN ID can be put in other places as described in [I-D.draft-li-apn-ipv6-encap], e.g., IPv6 extension header.

4. Overview

For simplicity, this document use APN segment as an illustration example, see Figure 1.

APN Seg: APN segment = LOC:FUNCT(APN ID): ARG

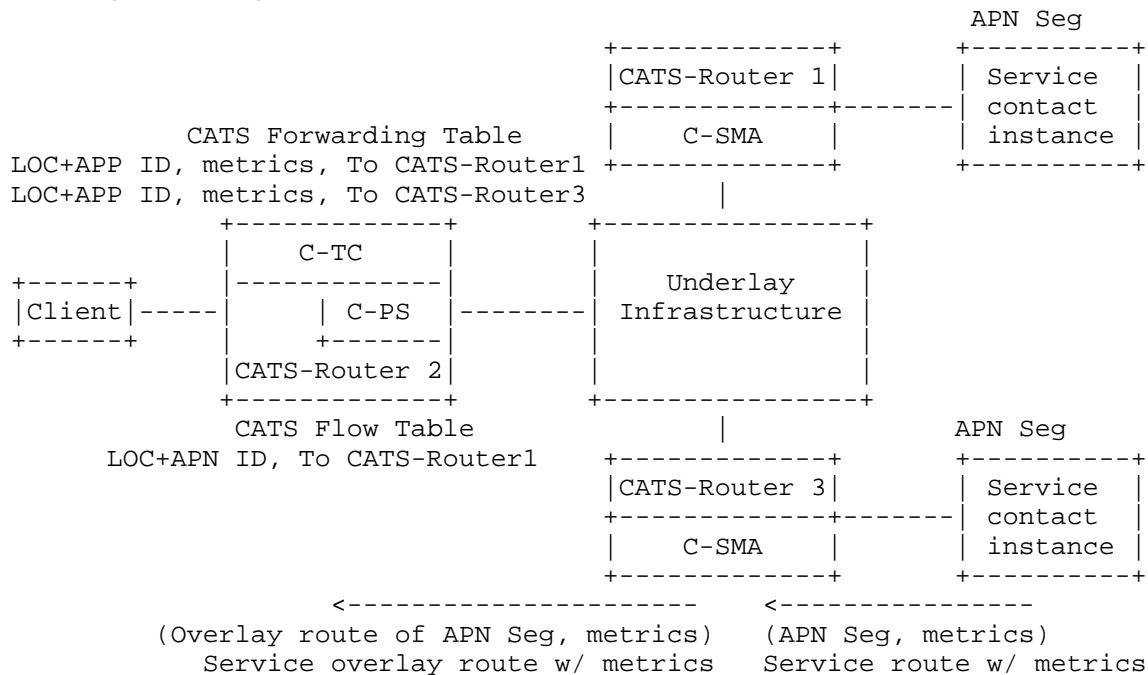


Figure 1: Architecture Overview

4.1. Realization of CATS Framework Components

The LOC + application group ID in the APN segment is used as CATS service identifier. The CATS overlay is established from the Ingress CATS-Router to an Egress CATS-Router. The CATS Traffic Classifier is running at the Ingress CATS-Router. Depending on the deployment, the CATS Path Selector, C-SMA and C-NMA can be centralized or distributed. CIS-ID is used to forward the packet to a specific CATS service contact instance. The choice of the CIS-ID is out of scope.

4.2. Realization of CATS Framework Workflow

4.2.1. Service Announcement

Locator is routable and leads to service contact instances. The locator is announced in anycast. The locator and application group ID may be learned by client using a rendezvous service (DNS, for example). The user group ID may be learned through a protocol between the client and server. The detailed rendezvous procedure is out of scope.

4.2.2. Metric Distribution

As described in CATS framework, the metrics needs to be distributed along with the CS-ID route. The detailed control plane solution depends on the deployment model (distributed, centralized or hybrid) and is out of scope of this document. A sample procedure using distributed model is provided to illustrate the core process, see Figure 1.

The C-SMA running as stand alone component at each service contact instance. In addition to the route announcement of LOC, C-SMA also distributes the application group ID and the computing metric to the Egress. The egress then pass the [(LOC, application group ID), computing metric] to the ingress node. The protocol extension used to carry the information is out of scope.

4.2.3. Service Request Processing

The service request packet has the destination address set to APN segment. The C-TC component in Ingress uses LOC + application group ID to match the service, finding the optimal service contact instance. Then the best path to the corresponding service instance is selected by C-PS. The packet is encapsulated and forwarded using the overlay path. The Egress decapsulates the packet then forward it to the service contact instance.

4.2.4. Service Instance Affinity

As per [I-D.draft-ldbc-cats-framework], the packets belongs to the same flow should goes to the same service contact instance. This document defines flow identifier as the tuple (locator, user group ID, app group ID). By assign the user group ID and app group ID, the service can archive customizable flow affinity without relying on traditional (IP, port) tuple.

5. APN for Real Locator

[I-D.draft-shi-cats-with-real-locator] defines real locator for CATS. When implementing solution of APN for CATS, the real locator can be carried in APN Para. Other procedure is the same as defined in [I-D.draft-shi-cats-with-real-locator].

6. Security Considerations

This document does not introduce any new security considerations.

7. IANA Considerations

TBD.

8. Acknowledgements

Much thanks to Hang Shi and Xia Chen for their contributions to the document.

9. References

9.1. Normative References

- [I-D.draft-li-apn-header]
Li, Z., Peng, S., and S. Zhang, "Application-aware Networking (APN) Header", Work in Progress, Internet-Draft, draft-li-apn-header-04, 12 April 2023, <<https://datatracker.ietf.org/doc/html/draft-li-apn-header-04>>.
- [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>>.
- [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>>.
- [I-D.draft-li-apn-ipv6-encap]
Li, Z., Peng, S., and C. Xie, "Application-aware IPv6 Networking (APN6) Encapsulation", Work in Progress, Internet-Draft, draft-li-apn-ipv6-encap-07, 10 July 2023, <<https://datatracker.ietf.org/doc/html/draft-li-apn-ipv6-encap-07>>.

[I-D.draft-shi-cats-with-real-locator]

Shi, H., Chen, X., and Z. Li, "CATS based on Real Locator", Work in Progress, Internet-Draft, draft-shi-cats-with-real-locator-00, 23 October 2023, <<https://datatracker.ietf.org/doc/html/draft-shi-cats-with-real-locator-00>>.

9.2. Informative References

[I-D.ietf-cats-usecases-requirements]

Yao, K., Trossen, D., Boucadair, M., Contreras, L. M., Shi, H., Li, Y., and S. Zhang, "Computing-Aware Traffic Steering (CATS) Problem Statement, Use Cases, and Requirements", Work in Progress, Internet-Draft, draft-ietf-cats-usecases-requirements-00, 24 July 2023, <<https://datatracker.ietf.org/doc/html/draft-ietf-cats-usecases-requirements-00>>.

[I-D.draft-ldbc-cats-framework]

Li, C., Du, Z., Boucadair, M., Contreras, L. M., Drake, J., Huang, D., and G. S. Mishra, "A Framework for Computing-Aware Traffic Steering (CATS)", Work in Progress, Internet-Draft, draft-ldbc-cats-framework-03, 4 August 2023, <<https://datatracker.ietf.org/doc/html/draft-ldbc-cats-framework-03>>.

Authors' Addresses

Zhenbin Li
Huawei
Beijing
Email: lizhenbin@huawei.com

Jianwei Mao
Huawei
Beiqing Road
Beijing
China
Email: maojianwei@huawei.com