

CATS WG
Internet-Draft
Intended status: Standards Track
Expires: 4 September 2025

CJ. Bernardos
UC3M
A. Mourad
InterDigital
3 March 2025

Terminal Mobility-Enabled Computing Aware Traffic Steering using IP
address anchoring
draft-bernardos-cats-anchoring-terminal-mobility-00

Abstract

The IETF CATS WG addresses the problem of how the network infrastructure can steer traffic between clients of a service and sites offering the service, considering both network metrics (such as bandwidth and latency), and compute metrics (such as processing, storage capabilities, and capacity).

This document defines new extensions and procedures for a terminal connected to a network infrastructure, to benefit from transparent mobility management adapting to specific connectivity and computing requirements, so traffic is always steered to an instance meeting both requirements. Both CATS-aware and -unaware terminals are considered.

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 4 September 2025.

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 and Problem Statement	2
1.1. Use case scenario	2
1.2. Problem statement	5
2. Terminology	5
3. Enabling terminal mobility with IP anchor mobility for CATS	5
3.1. OPTION A: terminal-aided	6
3.2. OPTION B: ECR-aided	11
3.3. OPTION C: CATS controller-aided	11
4. IANA Considerations	11
5. Security Considerations	11
6. Acknowledgments	11
7. Informative References	11
Authors' Addresses	12

1. Introduction and Problem Statement

1.1. Use case scenario

Let's consider a possible use case scenario, just for the sake of illustrating the scenario. Several nodes (UEs in this example) are acting as sensors in an Integrated Sensing and Communications (ISAC) case. The sensors generate/collect sensing data that needs to be processed timely and appropriately to generate an accurate sensing result. Part of this service is executed in the network infrastructure, posing some requirements on the connectivity (e.g., delay between the terminals and the node where the service is executed on the network infrastructure) and computing resources (e.g., capabilities to render the XR video within a certain latency budget). Within the network domain where the terminals are connected to there are multiple sites capable of hosting the service, each with potentially different connectivity and computing characteristics. Figure 1 shows an exemplary scenario. Considering the connectivity and computing latencies (just as an example of metrics), the best service site is #n-1 in the example used in the Figure.

Note that this is just an example, other services would also benefit from compute and connectivity traffic steering. For the sake of having a simpler service, we can also consider an AR/VR/XR service where a terminal connected to the network needs to instantiate a service in the network to aid in the AR/VR/XR service by providing computing capabilities with latency constraints.

Note on terminology. In this document we use the old terminology in which by ICR we mean Ingress CATS-Forwarder [I-D.ietf-cats-framework], and by ECR we mean Egress CATS-Forwarder.

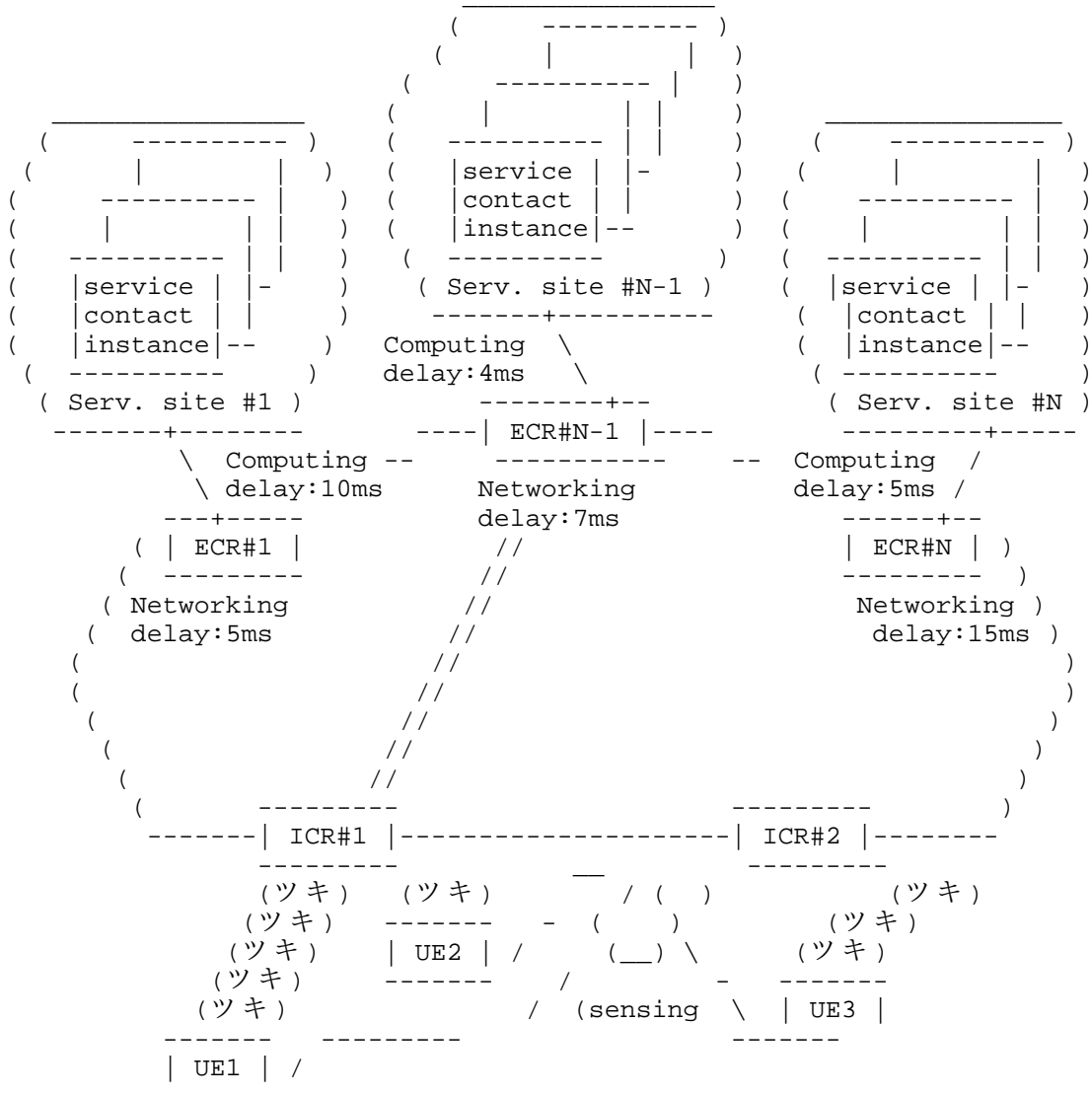


Figure 1: Exemplary scenario

1.2. Problem statement

The main problem that this document tries to address is the following. Networking systems do not have mechanisms yet with capabilities to support to dynamically change of point of attachment of a node which is consuming a service instantiated in the network, with the traffic being steered in accordance with the dynamically changing connectivity and computing conditions.

Based on the former, this document proposes solutions to enable the network to react and adapt to the change in connectivity and computing conditions caused by a terminal change of point of attachment, which might also trigger optimal service migration based on service-specific IP anchor mobility. In particular, this document addresses the following questions: (i) what mechanisms does the network need to implement to facilitate CATS-enabled terminal mobility, leveraging the architecture defined in [I-D.bernardos-cats-ip-address-anchoring], so the requirements of the service in terms of computing and networking are simultaneously fulfilled?, (ii) how to steer traffic between sensing terminals and the sensing service computing site, after sensor mobility, and how to re-evaluate --upon sensor mobility-- if service relocation is also needed, in a transparent manner to the sensor, by using IP anchor mobility?

2. Terminology

The following terms used in this document are defined by the IETF:

ECR: Egress CATS router. This refers to the Egress CATS-Forwarder as defined in [I-D.ietf-cats-framework].

ICR: Ingress CATS router. This refers to the Ingress CATS-Forwarder as defined in [I-D.ietf-cats-framework].

3. Enabling terminal mobility with IP anchor mobility for CATS

We describe next an example of operation and signaling for the network to perform terminal mobility. Three different options are described next, for variations (OPTIONS) of the procedures: terminal initiated, ECR-initiated and CATS-controller initiated. In addition to the functionality defined in [I-D.bernardos-cats-ip-address-anchoring] and [I-D.bernardos-cats-anchoring-service-mobility], this documents defines a new functionality:

- * Terminal mobility: it deals with the procedures required to (i) detect or predict a change of the current conditions due to a change in the point of attachment, jointly considering computing and networking, requiring of a service mobility operation; (ii) deciding whether a change of service instance location is needed, and, if so, selecting the best target service instance location and triggering the service migration, and (iii) ensuring that traffic is steered to the new point of attachment of the terminal. For example, a terminal or ICR might use this functionality to predict terminal mobility and prepare the network in advance, in cooperation with other CATS agents running at the current ICR, ECR and or service site. It is also used to perform the actual terminal mobility.

3.1. OPTION A: terminal-aided

Next, we assume terminal mobility is triggered by a CATS-aware terminal. By having a CATS agent running on the terminal, it can perform different monitoring actions to predict or detect the need to move from one point of attachment or another, and also the potential need to migrate a service from one site to another. This CATS agent might, for example, interact with other CATS agents deployed on ICRs, ECRs and service sites.

In the following we describe a terminal mobility procedure for CATS, initiated by a CATS-aware terminal. Using the sensing terminal support, the network infrastructure is capable of steering the traffic to the new location of the sensing terminal, and, if required, select a target service instance meeting the connectivity and computing requirements of the service. Both the mobility of the sensing terminal, and the potential service migration, are performed with signaling procedures transparent for the sensing terminal. Extensions and new behavior are highlighted. Note that variations are possible over this exemplary signaling diagram.

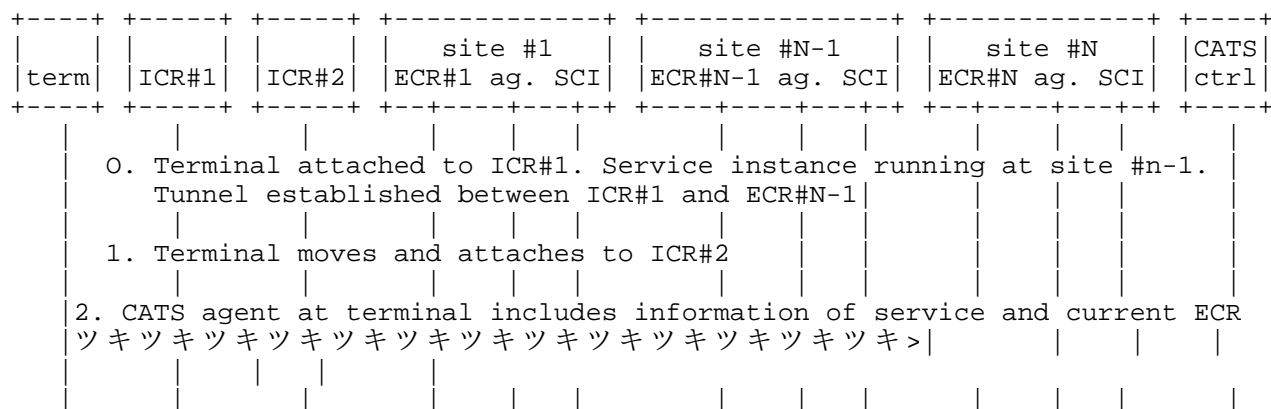




Figure 2 shows the message sequence chart of terminal mobility for CATS, initiated by a CATS-aware terminal (OPTION A), which is explained next:

0. A service/app has been already instantiated on a service site (for example, by using the mechanisms specified in [I-D.bernardos-cats-ip-address-anchoring]). In this example, the site where the service/app consumed by the terminal is currently instantiated is site #n-1. The terminal is connected to ICR #1 and there is a service tunnel between ICR#1 and ECR#N-1. This service/app requires some functionality to be run on the network infrastructure (e.g., an AR/VR/XR service). This service has specific requirements in terms of both connectivity and computing (CATS requirements).
1. The sensing terminal moves to a new ICR, changing its point of attachment from ICR#1 to ICR#2.
2. In this OPTION A, the sensing terminal is CATS-aware, and so it can provide additional information to the new ICR, as for example the ECR that is currently being used for the active service flow(s). How this is made is outside the scope of this document, but possible mechanisms include using extensions at IPv6 Neighbor Discovery (ND) level (e.g., in a Neighbor or Router Solicitation, NS or RS), or even at layer-2.
3. The current/new ICR (ICR#2 in this example) decides whether a change of sensing service site (and serving ECR) is needed or not, based on the compute and networking conditions at the new location of the terminal, and the requirements of the service. There are two possible outcomes: 1) the sensing service needs to be migrated because at the current location the required conditions cannot be met; 2) the service does not need to be migrated, and just the traffic needs to be steered to the new ICR from the currently serving ECR.
4. (SERVICE NEEDS TO BE MIGRATED:) The ICR sends a query to all (but currently used) ECRs of the domain, or a subset selected based on the location of the ICR. This query may include the following parameters:
 - i. Service ID: an identifier of the service requested by the terminal. This allows to check if the service can be instantiated or it is already instantiated.

- ii. Terminal ID: an identifier of the terminal requesting the service. This is useful for example for affinity purposes. It might not include information that can be used to identify the user.
 - iii. ICR ID: identifier of the requesting ICR.
 - iv. CATS requirements: list of requirements, e.g., connectivity and computing requirements.
5. Each ECR, possibly after checking with the CATS agent of the site(s) it provides connectivity, responds, including the following information:
- i. Service ID.
 - ii. Terminal ID.
 - iii. ECR ID: identifier of the ECR sending the response.
 - iv. CATS conditions: how the site meets each of the requirements included in the request.
 - v. (Optional): URI to get to the service instance.

A CATS agent at a site might be collocated with the ECR. Examples of a CATS agent at a site are network controllers or orchestrators at the site. Note that the way a CATS agent at an ECR may interact with the CATS agent of the site is out of the scope of this document. Examples include using monitoring and telemetry interfaces with an orchestrator managing the site.

6. Based on the received responses, and considering both networking and computing metrics and policies, the ICR selects an ECR (#n).
7. (SERVICE DOES NOT NEED TO BE MIGRATED:) The ICR requests the proposed/selected ECR to establish a traffic steering session with it, sending a CATS request. This request includes the same information that was included in the CATS query (to facilitate stateless operation of the ECRs while being queried), plus the following additional parameters:
- * ECR prefix: currently in use IP prefix IP to the terminal to reach the service instance.
 - * Lifetime: requested duration for the association between the ICR and the ECR.

8. The selected ECR responds back with an acknowledgement, including the following information:
 - i. Service ID.
 - ii. Terminal ID.
 - iii. ECR ID: identifier of the ECR sending the response.
 - iv. CATS conditions: how the site meets each of the requirements included in the request.
 - v. IP prefix assigned for the terminal to use to reach the service instance. It should match the one included in the request.
 - vi. Lifetime: granted duration of the association between the ICR and the ECR.
9. An IP tunnel is established between the ICR and the new ECR. Optionally (not shown in the figure), the ICR might send a CATS request with zero lifetime to the old ECR to remove the old tunnel.
10. The ICR sends a CATS request message to the old ECR so it can remove the state associated to the sensing terminal and service. The old ECR sends a CATS ACK to the new ICR.
11. The old ECR (ECR#n-1 in this example) also sends an unsolicited CATS response to the old ICR (ICR#1 in this example), so it can remove the associated state.
12. The previous message triggers the service migration (from site #n to site #n-1). The specific mechanism is out of the scope of this document. Note that some preparation/migration steps might be conducted in parallel (e.g., after messages #1 and #3) to accelerate the process, making this step just the final trigger for the service migration. At site #n, the prefix used by the terminal for accessing the service is configured to be used by migrated instance. This might require routing updates to be performed in the site, potentially controlled by a CATS agent running in the site.
13. Traffic of the service for this terminal is steered using the IP tunnel.

3.2. OPTION B: ECR-aided

TBD.

3.3. OPTION C: CATS controller-aided

TBD.

4. IANA Considerations

TBD.

5. Security Considerations

TBD.

6. Acknowledgments

The work of Carlos J. Bernardos in this document has been partially supported by the Horizon Europe MultiX (Grant 101192521), DISCO6G-CM (TEC-2024/COM-360) and UNICO I+D 6G-DATADRIVEN projects.

7. Informative References

[I-D.bernardos-cats-anchoring-service-mobility]

Bernardos, C. J. and A. Mourad, "Service Mobility-Enabled Computing Aware Traffic Steering using IP address anchoring", Work in Progress, Internet-Draft, draft-bernardos-cats-anchoring-service-mobility-03, 3 March 2025, <<https://datatracker.ietf.org/doc/html/draft-bernardos-cats-anchoring-service-mobility-03>>.

[I-D.bernardos-cats-ip-address-anchoring]

Bernardos, C. J. and A. Mourad, "Computing Aware Traffic Steering using IP address anchoring", Work in Progress, Internet-Draft, draft-bernardos-cats-ip-address-anchoring-03, 26 February 2025, <<https://datatracker.ietf.org/doc/html/draft-bernardos-cats-ip-address-anchoring-03>>.

[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-05, 10 February 2025, <<https://datatracker.ietf.org/doc/html/draft-ietf-cats-framework-05>>.

Authors' Addresses

Carlos J. Bernardos
Universidad Carlos III de Madrid
Av. Universidad, 30
28911 Leganes, Madrid
Spain
Phone: +34 91624 6236
Email: cjbc@it.uc3m.es
URI: <http://www.it.uc3m.es/cjbc/>

Alain Mourad
InterDigital Europe
Email: Alain.Mourad@InterDigital.com
URI: <http://www.InterDigital.com/>