CATS Working Group
Internet-Draft
Intended status: Informational
Expires: November 21, 2025

Jaehwoon Lee
Dongguk University

May 22, 2025

Network-based mobility management in CATS network environment
draft-jaehwoon-cats-mobility-03

Abstract

Computing-Aware Traffic Steering (CATS) network architecture is to choose the best edge computing server by considering both the network environment and available computing/storage resources of the edge computing server. This draft describes the mechanism in which service continuity is provided even when the client moves and connects to a new ingress CATS-Router by using the PMIPv6-based mobility management method in the CATS-based edge computing networking environment.

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 November 21, 2025.

Table of Contents

1. Introduction

Cloud computing provides powerful computing and nearly unlimited storage resources to client devices connected over the Internet. However, if the number of clients, such as Internet of Things (IoT) devices is quite large, the amount of traffic exchanged between clients and the cloud computing server is also high and it can cause congestion over the Internet. When congestion occurs on the path between a client and the cloud computing server, the client transmitting service request may experience long response time in receiving the result of the service request, or the service request itself may be lost.

In edge computing, even though edge computing server provides smaller computing and storage resources compared to the cloud computing server, multiple number of edge computing servers can be located near client devices and the client sending service request can benefit from shorter response time. In the edge computing environment, one way for a client to find a suitable edge computing server is to choose the nearest edge server based on the location of the client inferred from the client's source IP address. Another way is to choose one of the several edge servers by utilizing the round-robin method. However, in such cases, either the available resource in the chosen server can be insufficient or poor network condition between the client and the chosen edge server may result in long response time for the service request of the client.

IETF CATS working group tries to standardize the mechanism to choose the best edge computing server by considering both the networking environment and available computing/storage resources of the edge computing server[1]. Here, a service is represented by an CATS Service ID (CS-ID). Assume that there is a client trying to

receive a service provided by a specific service instance. In this case, ingress CATS-Router acts as a gateway for the client. In addition, egress CATS-Router is connected to the edge computing server in which specific service instance is installed. Assume that there are N edge servers providing the same specific service. Moreover, we assumet that different edge servers are connected to different egress CATS-Routers. The client transmits a service request message with CS-ID as a destination IP address. Ingress CATS-Router chooses the best egress CATS-Router by using the combination of the network metric such as delay, and computing metric such as available computing/storage resource of edge servers. The ingress CATS-Router transmits the service request sent by the client to the chosen egress CATS-Router. After which egress CATS-Router transmits the service request to the service instance in the edge computing server. The result of the service request is in turn transmitted from the edge server to the client through the egress CATS-Router and the ingress CATS-Router.

When a client transmits a service request and then moves to another network before receiving the service result, the client can no longer receive the result of the service request. When the client moves and connects to a new ingress CATS-Router, host-based mobility management method such as Mobile IPv6 (MIPv6) can be used to maintain end-to-end connectivity[2]. In this case, the destination IP address of the service request message sent by the client is the CS-ID. This means that the new ingress CATS-Router cannot know the address of the egress CATS-Router connected to the edge server providing service to the client. Therefore, host-based mobility management cannot be used in the CATS networking environment. Network-based mobility management mechanism such as Proxy MIPv6 (PMIPv6) can be used in the CATS networking environment if the new ingress CATS-Router knows the address of the egress CATS-Router connected to the edge server providing service to the client[3]. In this case, service continuity is ensured for the client. However, new ingress CATS-Router cannot know the IP address of the egress CATS-Router only using the address information of the IP packet sent by the client. The reason is that the destination address of the IP packet indicates not the same egress CATS-Router that the client are connected but the same specific service.

As mentioned above, when moving from one network to another while the connection between the client and the edge server is established, if the transaction between the client and the edge server needs to continue for a considerable period of time, the current edge server to which the client is connected may not be the optimal server when considering both network resources and computing resources. In this case, the new ingress CATS-Router to which the client is connected needs to select a new optimal edge server considering both network resources and computing resources, and

transmit the client’s current state information from the existing edge server to the new edge server, so that the client can continue to receive computing service from the new optimal edge server.

This draft describes the mechanism in which service continuity is provided even when the client moves and connects to a new ingress CATS-Router by using the PMIPv6-based mobility management method in the CATS-based edge computing networking environment. Moreover, This draft, also describes a technique to enable a new ingress CATS-Router to select a new edge server suitable for the client’s service request when a client moves while establishing a connection with an edge server and receiving service responses, and to enable the client to continuously receive services from the new edge server.

2. Conventions and Terminology

2.1. Conventions

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL","SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [4].

2.2 Terminology

TBD.

3. Protocol Operation

When a client moves from an ingress CATS-Router to another ingress CATS-Router before receiving all the service results, either proactive method of reactive method can be utilized to provide service continuity.

Fig. 1 shows the message exchange procedure to provide service continuity proactively when a client moves to another network in CATS networking environment. If the client transmits service request message with CS-ID as a destination IP address, an ingress CATS-Router (that is, old ingress CATS-Router) chooses the best egress CATS-Router by using the combination of the network metric and computing metric. The old ingress CATS-Router transmits the service request to the chosen egress CATS-Router. The egress CATS-Router transmits the service request message to the corresponding service instance in the edge computing server. When the old ingress CATS-Router detects the movement of the client before completing transmission of all service results, it transmits the CATS mobility notification message including the addresses of the client and the chosen egress CATS-Router to one or more candidate new ingress CATS-Routers that client may connect to. The format of the CATS

mobility notification message is defined in Section 4.1. Here, how the old ingress CATS-Router can know the movement of the client is out of scope. One method is to use the signal strength of the client. Moreover, how the old ingress CATS-Router can know which is the new ingress CATS-Router that the client moves and connects to is TBD. One method is for the old ingress CATS-Router to broadcast the CATS mobility notification message to neighbor ingress CATS-Routers. Another method is to find some candidate ingress CATS-Routers by using the GPS information of the client. When the client moves and connects to a new ingress CATS-Router, the new ingress CATS-Router transmits the CATS mobility indication message having the IP address of the client to the old ingress CATS-Router and establishes the tunnnel with the old ingress CATS-Router. The format of the CATS mobility indication message is defined in Section 4.2. The old ingress CATS-Router having received the CATS mobility indication message also establishes the tunnel with the new ingress CATS-Router.

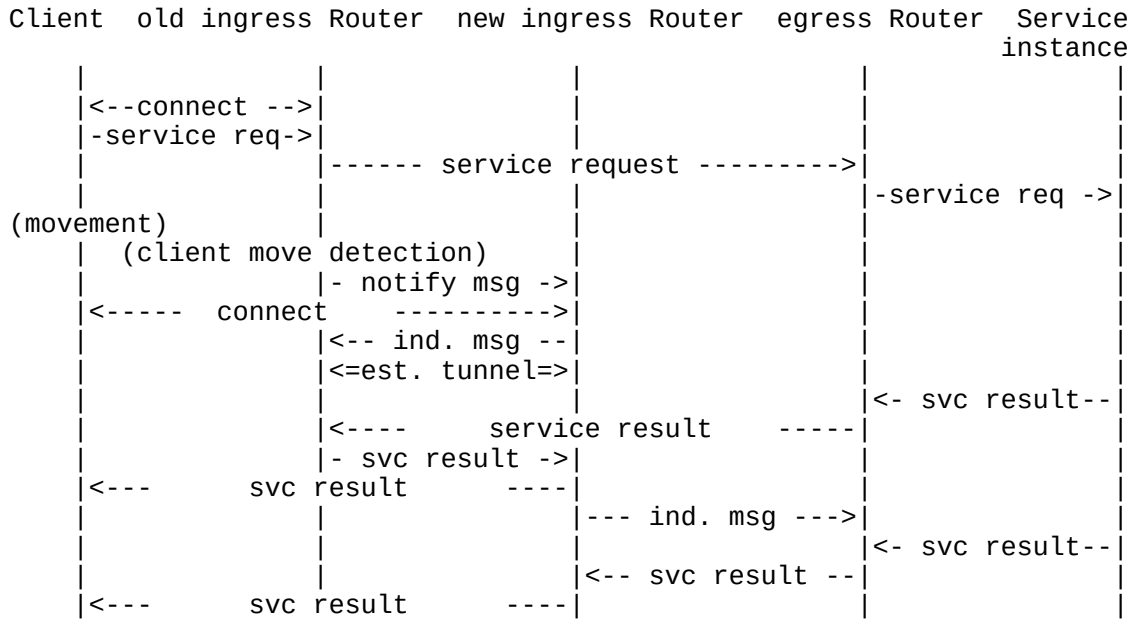


Figure 1: Message exchange procecure - proactive method

Moreover, the new ingress CATS-Router transmits the CATS mobility indication message having the client’s IP address and the IP address of the old ingress CATS-Router to the egress CATS-Router. From now on, the old ingress CATS-Router and the egress CATS-Router can transmit all services results to the client through the new ingress CATS-Router.

Fig. 2 shows the message exchange procedure to provide service continuity reactively to the client. If the client moves and connects to a new ingress CATS-Router, the new ingress CATS-Router transmits the CATS mobility request message including the IP address of the client to the old ingress CATS-Router. The format of the CATS mobility request message is defined in Section 4.1. Here, how the new ingress CATS-Router can know the address information of the old ingress CATS-Router is TBD. One method is to use a location server. When a client connects to an old ingres CATS-Router, the old CATS-Router store the IP and link layer addresses of the client and the IP address information of the egress CATS-Router that the service request of the client is transmitted in the location server. The information regarding the client can be removed just after all service results are transmitted to the client. When a client moves to a new ingress CATS-Router, then the new ingress CATS-Router can know whether the client is a new client or the client requiring service continuity by quering the information stored in the server.

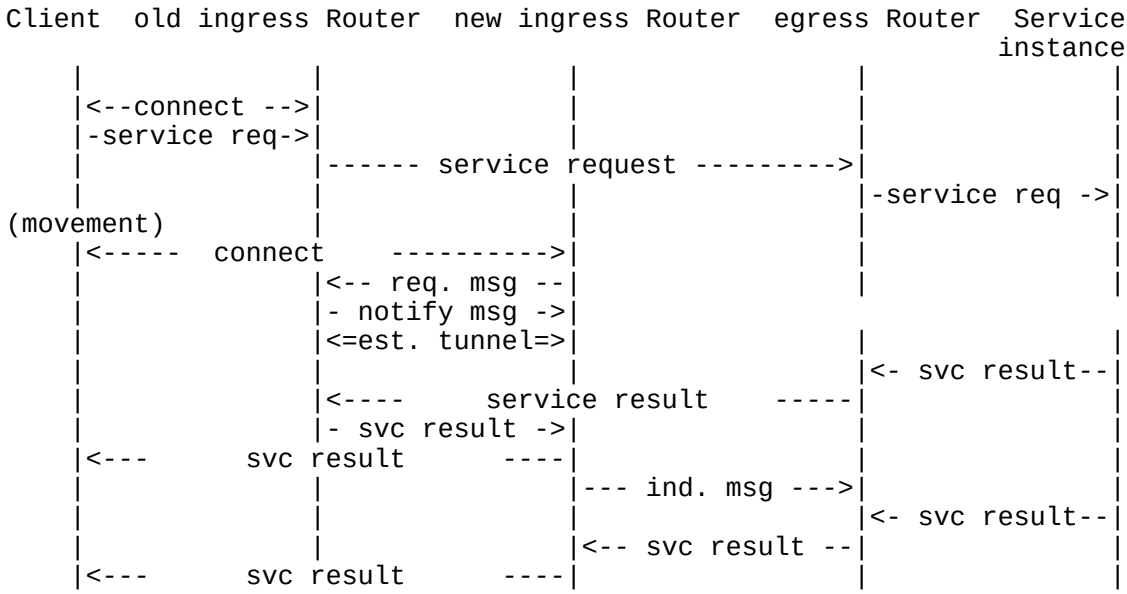
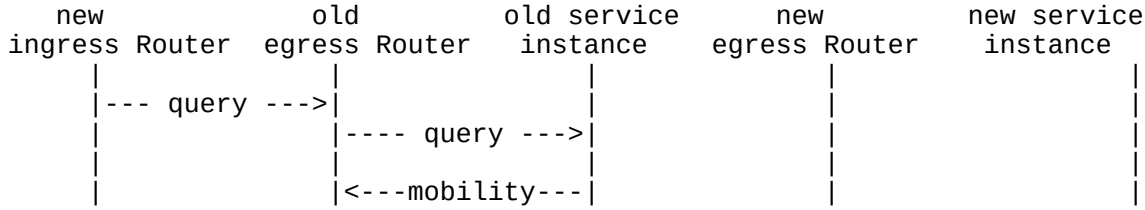


Figure 2: Message exchange procecure - reactive method

Another method is to assign a network address to CAT domain but different sub-network address is assigned to different ingress CATS-Router. For example, assume that 10.0.0.0/8 network address is assigned to a CATS domain. Here, 10.0.0.0/16 sub-network address is assigned to the old ingress CATS-Router and 10.0.1.0/16 sub-network address is assigned to the new ingress CATS-Router. Moreover, 10.0.0.1 IP address is assigned to the old ingress CATS-Router and 10.0.1.1 IP address is assigned to the new ingress CATS-Router. When

a client connects to the old ingress CATS-Router, the router advertises 10.0.0.0 network address by using the Router Advertisement message. If the client transmits DHCP request message requesting a new IP address, the router assigns one of the IP addresses belonging to 10.0.0.0/16 sub-network address. When the client moves and connects to the new ingress CATS-Router, the router advertises 10.0.0.0/8 network address by using the Router advertisement message. If the client transmits DHCP request message, then the router considers that the client is the newly connected client. Otherwise, the router can deduce the IP address of the old ingress CATS-Router by using the source IP address of the packet transmitted by the client. The old ingress CATS-Router having received CATS mobility request message transmits the CATS mobility notification message including the IP address of the egress CATS-Router to the new ingress CATS-Router and establishes the tunnel with the new ingress CATS-Router. Moreover the old ingress CATS-Router can inform the new ingress CATS-Router whether the client needs service continuity or not by using the notification message. The new ingress CATS-Router transmits the CATS mobility indication message to the old ingress CATS-Router and establishes the tunnel with old ingress CATS-Router. Moveover, it transmits the CATS mobility indication message to the egress CATS-Router. From now on, the old ingress CATS-Router and the egress CATS-Router can transmit all service results to the client through the new ingress CATS-Router.

Figure 3 shows the messages exchange procedure related to mobility between edge servers. The new ingress CATS-Router chooses a new optimal service instance using the service request information of the client and sends a query message to the old egress CATS-Router to ask whether service mobility is needed. The old egress CATS-Router receiving this message transmits the query message to the old edge server whether service mobility is needed. When the old edge server transmits the response message that service mobility is needed including the information related to the current service status, old egress CATS-Router transmits this information to the new egress CATS-Router. The new egress CATS-Router transmits this information to the new edge server and sends the notification information message to the new ingress CATS-Router that the mobility between service instances is complete. From now on, the client continues transaction the remaining service with new service instance connected to the new edge server.



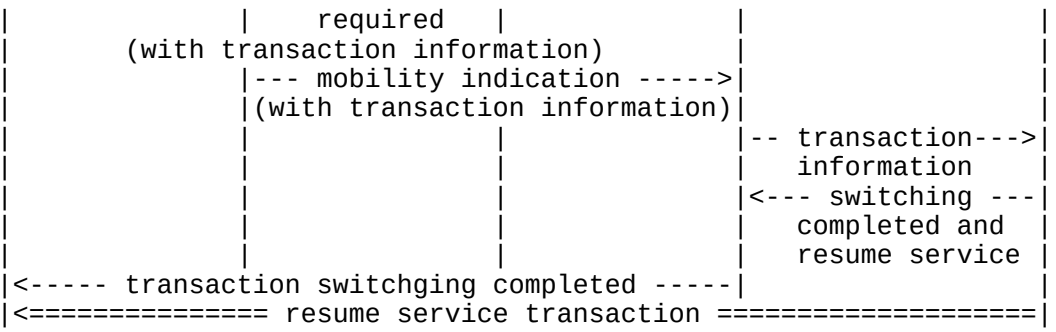
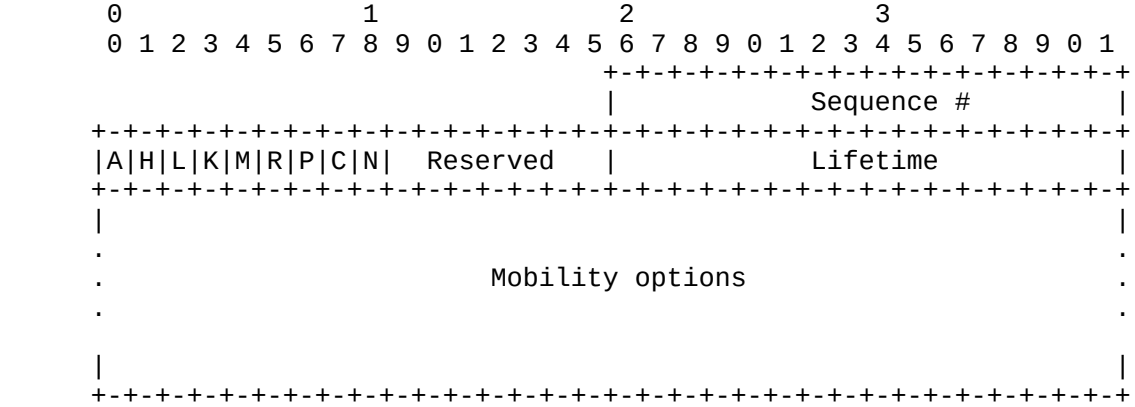


Figure 3: Transaction switching procecure

4. Message Formats

4.1 CATS mobility notification and request messages

In this draft, the proxy binding update message defined in the Proxy Mobile IPv6 protocol is used to define the CATS mobility notification and request messages [3]. The message format is as follows:



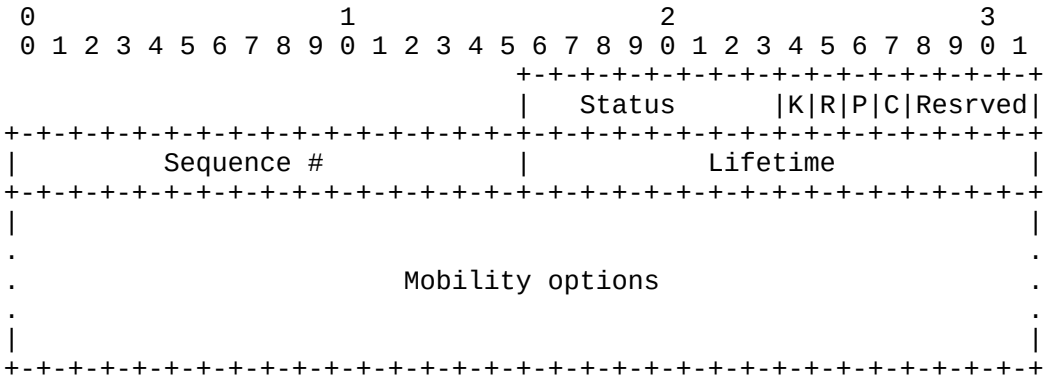
C: CATS flag. This bit must be set to 1 in the CATS environment.

N: The flag must be set to 0 for CATS mobility notification message must be set to 1 for CATS mobility request message.

The mobility option of the CATS notification message contains client node address option and CATS address option defined in Section 4.3. In this case, the address contained in the CATS address option is the egress CATS-Router address. Moreover, the mobility option of the CATS request message contains the client node address option.

4.2 CATS mobility indication message

In this draft, the proxy binding acknowledgment message defined in the Proxy Mobile IPv6 protocol is used to define the CATS mobility indication message [3]. The message format is as follows:

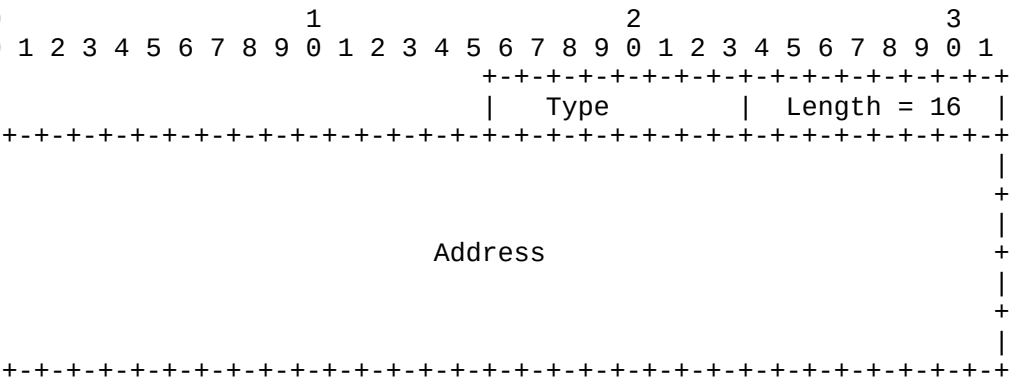


C: CATS flag. This bit must be set to 1 in the CATS environment.

When the message is transmitted from the new ingress CATS-Router to the old ingress CATS-Router, the client node address option is included in the mobility option. Moreover, when the message is transmitted from the new ingress CATS-Router the the egress CATS-Router, the client node address and CATS address options are included in the mobility options. In this case, the address included in the CATS address option is the old ingress CATS-Router.

4.3 Mobile node address and CATS address options

In this draft, the mobility option defined in the Mobile IPv6 protocol is used to define the client node address and CATS address options [2]. The option format is as follow:



Client node address option:
- Type : TBD
- The mobile node address is included in the Address field.

CATS address option:
- Type : TBD
- The CATS address is included in the address field.

5. Security Considerations

TBD

6. IANA Considerations

TBD

7. References

[1] C. Li, Z. Du, M, Boucadair, L. M. Contreras and J. Drake, "A Framework for Computing-Aware Traffic Steering (CATS)", draft-ietf-cats-framework-07 (work in progress), Apr. 30, 2025.

[2] D. Johnson, C. Perkins and J. Arkko, "Mobility Support in IPv6", IETF RFC 6275, July 2011.

[3] S. Gundavelli, K. Leung, V. Devarapalli, K. Chowdhury and B. Patil, "Proxy Mobile IPv6", IETF RFC 5213, Aug. 2008.

[4] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

Author’s Address

Jaehwoon Lee
Dongguk University
30, Pildong-ro 1 gil, Jung-gu
Seoul 04620, KOREA
Email: jaehwoon@dongguk.edu