IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 22 May 2024, accepted 22 June 2024, date of publication 24 June 2024, date of current version 1 July 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3418960

RESEARCH ARTICLE

Design of Computing-Aware Traffic Steering Architecture for 5G Mobile User Plane

MINH-NGOC TRAN, VAN-BINH DUONG, AND YOUNGHAN KIM, (Member, IEEE)
School of Electronic Engineering, Soongsil University, Seoul 06978, South Korea

Corresponding author: Younghan Kim (younghak@ssu.ac.kr)

This work was supported in part by the Institute of Information and Communications Technology Planning and Evaluation (IITP) Grants funded by Korean Government (MSIT), Development of Candidate Element Technology for Intelligent 6G Mobile Core Network, under Grant 2022-0-0101; and in part by the Development of High Available and High Performance 6G Cross Cloud Infrastructure Technology under Grant RS-2024-00398379.

ABSTRACT One of the critical targets of the fifth generation (5G) mobile network is supporting low latency and high reliable services at the network edge servers. Multiple service instances of the same service can be deployed on different geographically distributed edge sites for high availability purposes. However, the dynamic changes over time in edge servers' computing capabilities and underlay infrastructure network quality might affect service performance. To avoid service performance degradation, an efficient computing-aware traffic steering (CATS) solution between different service instance locations is required. Multiple advanced traffic steering algorithms have been proposed in different existing 5G traffic steering works. However, no available works addressed the possible 5G architecture and procedure changes to convert these CATS algorithms' decisions into 5G mobile user plane routing paths. This work aims to cover this research gap. We propose enhancing the current 5G architecture by applying the Mobile User Plane Controller (MUP-C) concept and using anycast address for multi-instance services. We discuss three implementation options to apply these enhancement features to 5G architecture: CATS Application Function (AF) traffic influence method, CATS MUP-C dummy User Plane Function (UPF) method, and CATS MUP-C non-UPF method. We implemented and evaluated the performance of our solutions against a 3GPP standard dynamic edge discovery method, which can also be used to support CATS. The results showed that our solutions had better Protocol Data Unit (PDU) session setup procedure efficiency for CATS than the standard method by reducing the setup latency from 30% to 50%.

INDEX TERMS 5G, computing-aware traffic steering, mobile user plane, segment routing.

I. INTRODUCTION

5g network promises to support high reliability, high availability, and low latency services. One of the key enabling features of 5G for this target is Mobile Edge Computing (MEC). According to a Cisco annual internet report from the last 5 years, this MEC deployment method has been well adopted by both enterprise and telecom companies over various edge computing application use cases. One particular service can be deployed closer to users with multiple service instances running over different MEC servers at different locations on the network edge.

Because of the different service instance location options, the 5G network needs to decide one of the available locations to steer an user equipment (UE)' service request. Meanwhile, different MEC servers' computing capabilities (e.g., CPU, memory resource) and 5G underlay network status (e.g., bandwidth, throughput) might frequently change over time. Insufficient resources or dramatic changes in computing status at the MEC node and sudden incidents happening at the underlay infrastructure network can cause detrimental effects on the service's user experience. Hence, 5G network need to consider this information to dynamically steer service traffic over different MEC locations to guarantee service experience for users. The problem of steering traffic between different service instance locations considering computing

The associate editor coordinating the review of this manuscript and approving it for publication was Irfan Ahmed.

and network information is defined as Computing-Aware Traffic Steering (CATS) by the CATS Working Group of The Internet Engineering Task Force (IETF) [1].

Although 5G traffic steering is a popular research topic, no existing work has considered the possible 5G architecture and procedure changes to support CATS. Previous works can be divided into two types of research problems. The first problem type is traffic steering at the 5G Radio Access Network level. Works targeting this problem propose solutions to steer traffic between different 5G base stations, or radio access technologies based on different network requirements of the deployed services. This problem type is out of scope of this article. The second problem type is traffic steering at the 5G data plane level. Works targeting this problem propose different kinds of advanced algorithms to steer traffic between different service instance locations. These algorithms can also be used for CATS. However, these works heavily focused on traffic steering algorithms. To implement and convert the traffic steering decisions of these algorithms into corresponding routing information at 5G data plane, changes to the 5G architecture and procedure might be required.

Figure 1 illustrates the scope of this article. We focus on 3 5G architecture design aspects to convert the traffic steering decision of the CATS algorithm into the corresponding 5G underlay datapath between UE and the decided data network (DN). The first aspect is the interaction between the entity that runs the CATS algorithm (we call this entity the CATS algorithm function in this article) and the 5G Core functions. This interaction interface defines how the traffic steering decision is provided to the 5G Control Plane. The second aspect is the interaction between the 5G control plane and the 5G data plane. The 5G control plane includes 5G core network functions. The 5G data plane includes many Uplink Classifier User Plane Functions (UPF UL-CL), PDU Session Anchor User Plane Functions (UPF PSA), and the underlay network infrastructure routers that create the 5G N3, N6, N9 interfaces between the UPFs. The control-data plane interaction defines how 5G core network functions configure the data path in the underlay network corresponding to the traffic steering decision. The third aspect is the location of the CATS Algorithm function in 5G architecture to achieve reasonable routing underlay data path setup latency. This function can be placed at the 5G control plane or the 5G data plane.

Current 5G architecture support for steering traffic between different service instance options was defined by the 3rd Generation Partnership Project (3GPP) as the Edge Discovery procedure in the technical specification document 23.548 [2]. The advanced traffic steering algorithm (CATS algorithm in this article) can be performed at a central Domain Name System (DNS) server or an Edge Application Server Discovery Function (EASDF). This entity resolves the UE's DNS query to decide the optimal service instance location. There are two options for setting up the underlay routing data path based on the traffic steering decision from the central DNS server/EASDF: dynamic session breakout, and pre-established session breakout. In the first option, UE's DNS query is routed via a central UPF PSA. Then, based on the response from the DNS server/EASDF, the 5G control plane reconfigures the PDU session to a local UPF PSA corresponding to the chosen service instance location. In the second option, the local UPF PSA for an UE request is pre-configured at the 5G control plane based on the current UE requesting location. UE's DNS query is routed to a local DNS resolver via the pre-configured local UPF PSA. The local DNS resolver communicates with the central DNS server/EASDF to reply the optimal service instance's address to UE.

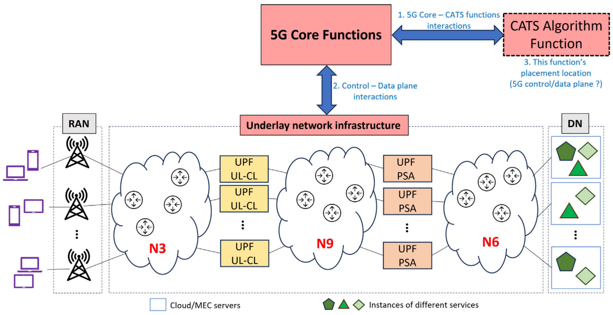
5G Core Functions
1. 5G Core – CATS functions interactions
CATS Algorithm Function
3. This function's placement location (5G control/data plane ?)
2. Control – Data plane interactions
Underlay network infrastructure
RAN
DN
UPF UL-CL
UPF UL-CL
UPF UL-CL
UPF PSA
UPF PSA
UPF PSA
N3
N9
N6
Cloud/MEC servers
Instances of different services

FIGURE 1. 3 5G architecture design aspects to support CATS.

However, neither of these options is optimal for supporting CATS. The pre-established session breakout option is unsuitable for dynamic computing and network status change. Meanwhile, the dynamic session breakout option requires establishing a PDU session via a central UPF PSA first before modifying it to a local UPF PSA later. This extra re-configuration disadvantage can be addressed by a solution that can directly establish the PDU session via the local UPF-PSA corresponding to the optimal service instance location chosen by the CATS algorithm.

In this work, we propose 5G architecture enhancements to support CATS. Our key contributions are as follows:

- We propose to add two enhancements to the current 5G architecture. First, we use the MUP-C concept [3] to directly convert PDU session information to an optimal user plane routing path based on CATS information. Second, we use anycast addresses to represent multiple-instance services. This enhancement can reduce PDU session setup time caused by DNS resolution.
- We propose three possible implementation options to implement these architecture changes. The first option is the CATS AF traffic influence method. In this option, the CATS algorithm function runs as a 5G AF at the control plane and influences 5G core functions to set up the corresponding UPF PSA of the decided service instance location. For the other two options, the CATS algorithm function runs inside the MUP-C at the data plane. The second option is the CATS MUP-C dummy-UPF method. In this option, the 5G core functions set up the PDU session using a special dummy UPF.

The MUP-C at the data plane converts this session information into a corresponding routing path based on the CATS algorithm function decision. The third option is the CATS MUP-C non-UPF method. UPF configuration is not required in this method. The 5G Session Management Function (SMF) self-generates the session information instead of obtaining it from the UPF setup procedure.

- We implemented our three proposed deployment options of 5G mobile user plane CATS by using 5G and software-defined networking (SDN) open-source tools. We evaluate our three deployment options and the 3GPP edge discovery procedure regarding PDU session setup latency and routing path throughput.

This study is organized as follows: Section II reviews 5G Traffic Steering related works and background about CATS and MUP-C. Section III introduces our three proposed 5G CATS architecture implementation options. Section IV presents the implementation details. Section V discusses our solution's performance results. Section VI concludes the paper.

II. BACKGROUND AND RELATED WORKS

A. COMPUTING-AWARE TRAFFIC STEERING

Computing-aware traffic steering is a problem defined by the IETF CATS working group. It targets the networking scenario where multiple instances of a service are deployed at different edge servers to achieve low latency and high availability. The computing resource availability and network conditions linked to edge servers can frequently change over time. Lack of resources can negatively affect service performance. The network should consider computing and network information to steer and load balance traffic dynamically to appropriate edge servers to avoid service performance degradation.

The IETF CATS working group defines a general CATS framework [4] to address this problem. Figure 2 describes the architecture of this framework. The CATS framework includes the following components: CATS Service Metrics Agent (C-SMA), CATS Network Metrics Agent (C-NMA), CATS-Forwarder, CATS Path Selector (C-PS), and Traffic Classifier (C-TC). C-SMA and C-NMA are the metric agents responsible for collecting computing and network information. The C-SMA can gather information related to the deployed services, such as service sites' computing resource status, service instance deployment information, and service instance serving queue information. Meanwhile, the C-NMA can gather any network information at the underlay network infrastructure, such as bandwidth and latency of different network paths linked to the service instances. The C-PS uses the collected information from C-SMAs and C-NMAs to decide the best service instance to solve a user request. The CATS-Forwarder is responsible for forwarding traffic based on the C-PS's decision. The C-TC ensures that packets bound to a specific contact instance are all forwarded along the same path. C-SMA and C-PS can be deployed as decentralized components alongside the CATS-Forwarder or a centralized component. Metrics distribution mechanisms and the CATS service instance selection algorithm are not included in the scope of the CATS framework. The CATS framework is open to any mechanisms or algorithms. It only defines the required components to enable CATS.

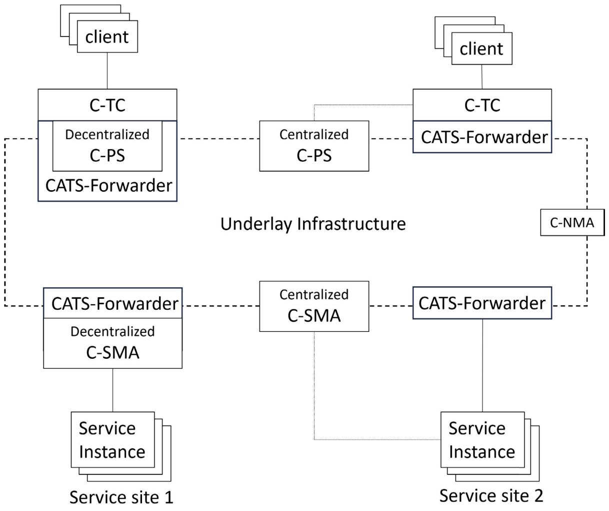

FIGURE 2. Computing-aware traffic steering (CATS) architecture.

However, the target of the current CATS framework design is the general Internet. Applying the CATS framework concept to the 5G network might require 5G architecture or procedure changes. This article proposes 3 different 5G CATS implementation options to address this research gap.

B. PREVIOUS 5G TRAFFIC STEERING WORKS

5G Traffic Steering is a popular research area. Based on the research target, previous 5G traffic steering works can be divided into two categories: 5G RAN traffic steering and 5G data plane traffic steering. In this sub-section, we briefly review the most recent works of these two categories and specify the remaining research gap that our work addresses.

The 5G RAN traffic steering category includes different types of traffic steering problems at the RAN level of the 5G network. Most 5G traffic steering works belong to this category. Optimal 5G base station/cell selection to serve UE traffic is one of the popular problems. In [5] and [6], different reinforcement learning-based Open RAN traffic steering frameworks were proposed to optimally assign a based station to UE. Another Open RAN solution was proposed in [7] to optimize RAN slicing and flow split distribution based on traffic prediction. The authors of [8] proposed a machine learning-based solution that can predict the optimal cell for UE connections based on mobility and network conditions. In [9], the authors proposed a federated learning solution that allows UE to make traffic steering decisions instead of base stations. Machine learning was used in [10] to forecast the quality of service that UE will get from the network to choose the optimal base station distributed unit. The work [11] discussed the performance of various machine learning techniques for this 5G RAN traffic steering problem. Meanwhile, another type of problem in the 5G RAN traffic steering category is traffic steering over different radio access networks (e.g., long-term evolution, new radio, wireless local-area network). Different optimization targets are considered to select the appropriate network, such as the quality of service constraint in [12] and [13] or throughput and energy efficiency in [14].

The 5G data plane traffic steering category includes a few works that aim to select the optimal data plane routing path and service instance location to serve UE requests. In [15], the authors created a graph that can capture network and edge application layer information. Then, they applied the Multi-objective Dijkstra algorithm to analyze the graph and select the routing path to the optimal edge server. In [16], the authors proposed an SDN-based solution to efficiently steer traffic to appropriate MECs based on UE mobility, wireless network, and MEC server resource information. The authors of [17] proposed a matrix-based dynamic shortest path selection algorithm to find the lowest delay cost and optimized bandwidth path to steer traffic to the MEC servers during vehicle mobility.

This study targets the 5G data plane traffic steering problem category. The previous works of this category have provided several advanced algorithms to find the optimal traffic steering routing path. These proposed algorithms can also be considered for the CATS problem because they also consider computing and network information. However, no work has considered the required 5G procedure or architecture changes to convert the optimal traffic steering decision from CATS algorithm into the real 5G underlay routing data path. We address this gap in our proposed solutions.

C. 3GPP'S 5G ARCHITECTURE SUPPORT FOR MULTIPLE-INSTANCE SERVICE TRAFFIC STEERING

3GPP provides technical baseline standards and specifications for telecom companies to build cellular, 5G, and future 6G networks. The problem of traffic steering between different service's serving instances has been addressed in 3GPP Technical Specification 23.548 [2]. Specifically, this specification refers to this problem as the Edge Discovery problem. The 3GPP document describes two types of 5G edge discovery procedures: dynamic session breakout and pre-established session breakout. In this sub-section, we briefly review these procedures and discuss their current limitations.

Figure 3 illustrates the 3GPP's edge discovery dynamic session breakout procedure. In this procedure, the user plane routing data path for the UE request is only configured after the 5G control plane receives the chosen service instance from the DNS server. First, when UE requests the service, the 5G SMF chooses a suitable EASDF instance to serve the UE's DNS query. Then, the SMF configures a UPF PSA for DNS communication between the EASDF and the DNS server. The EASDF then receives the DNS query from UE and asks the DNS server to resolve it. The DNS server can respond with one or a list of candidate service instance addresses. If the DNS server responds with a single service instance address, we can suppose that the CATS algorithm runs at the DNS server. Otherwise, the CATS algorithm runs at the EASDF to select an optimal service instance from the candidate service instance list. The EASDF then forwards the decided service instance information to the 5G SMF. The 5G SMF then selects and configures the suitable UPF ULCL and local UPF PSA to steer the UE request toward the chosen service instance. Finally, the DNS response is forwarded to UE, and UE traffic can be routed to the optimal service instance.

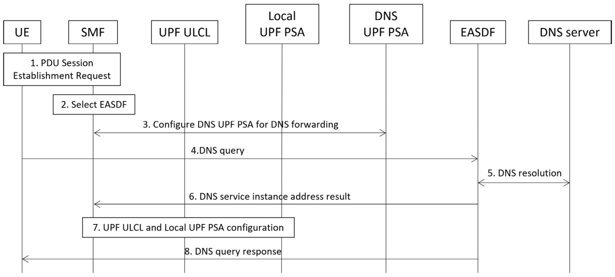
UE
SMF
UPF ULCL
Local UPF PSA
DNS UPF PSA
EASDF
DNS server
1. PDU Session Establishment Request
2. Select EASDF
3. Configure DNS UPF PSA for DNS forwarding
4.DNS query
5. DNS resolution
6. DNS service instance address result
7. UPF ULCL and Local UPF PSA configuration
8. DNS query response

FIGURE 3. 3GPP's dynamic session breakout edge discovery traffic steering.

The 3GPP edge discovery dynamic session breakout procedure is enough to support user plane routing data path configuration based on CATS algorithm. However, setting up the routing path requires configuring a corresponding UPF PSA of the DNS server every time a DNS resolution is required. This extra UPF configuration might introduce latency overhead to traffic steering configuration. In our 3 5G CATS deployment options proposed in this article, we address this limitation using the AF traffic influence procedure or the MUP controller concept to convert the CATS algorithm decision. Besides, we use anycast address instead of the traditional DNS resolution method to represent the service. The details are described in later sections.

Figure 4 illustrates the 3GPP's edge discovery pre-established session breakout procedure. In this procedure, the user plane routing data path for the UE request is pre-configured without dependency on the DNS response. In contrast with the previously mentioned method, this one is a static method. The optimal service instance to serve the UE request might be decided in advance. For example, a specific service instance might be pre-configured to serve any UE requests from a pre-defined location area. Service provider might use an 5G AF to inform the pre-configured traffic steering rules to the 5G control plane. First, when UE requests the service, the 5G SMF configures the pre-configured UPF ULCL and local UPF PSA. Then, the UE DNS query is forwarded via the local UPF PSA and resolved by the local DNS resolver and server. The DNS response is only used for replying the pre-determined service instance address to UE.

The 3GPP edge discovery pre-established session breakout procedure is a static traffic steering configuration method that does not fit the scope of this document. CATS requires

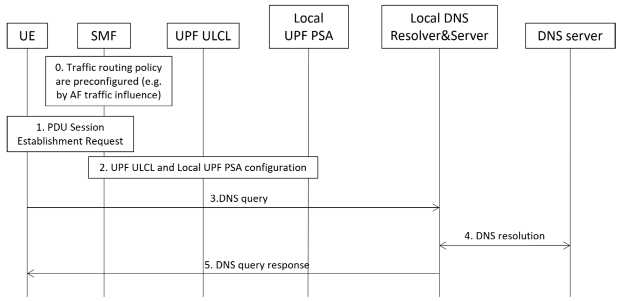
UE
SMF
UPF ULCL
Local UPF PSA
Local DNS Resolver&Server
DNS server
0. Traffic routing policy are preconfigured (e.g. by AF traffic influence)
1. PDU Session Establishment Request
2. UPF ULCL and Local UPF PSA configuration
3.DNS query
4. DNS resolution
5. DNS query response

FIGURE 4. 3GPP's pre-established session breakout edge discovery traffic steering.

a traffic steering method that might dynamically configure different service instances to different UE requests of the same service based on computing and network information.

D. THE MUP CONTROLLER CONCEPT

In the current 5G network, a PDU session is required to route UE traffic to DN. For each PDU session, 5G SMF configures a specific UPF PSA as the anchor node. UE traffic packets are transmitted in the 5G underlay network via a static General packet radio service Tunneling Protocol (GTP) tunnel between gNodeB and the UPF PSA. According to several IETF discussions about mobile user plane evolution ([18], [19]), the UPF anchor factor makes it difficult for the operator to optimize the user plane data path. UPF re-configuration procedure is required whenever the UE data path changes (e.g., due to UE mobility, edge service relocation, etc.). Besides, the routing path over UPF might not be the optimal direct path between UE and DN, according to a recent demonstration by SoftBank [20]. Meanwhile, alternative underlay network protocol options that support dynamic traffic engineering capability, such as Segment Routing (SR), and Locator ID Separation Protocol (LISP), can be utilized to optimize the 5G user plane [21].

Recently, a new mobile user plane architecture approach was proposed in the IETF Distributed Mobility Management working group to address this issue [3]. This study refers to this approach as the DMM MUP architecture. This approach introduces the MUP-C that bridges the 5G control plane and the user plane's network protocols, such as SRv6 or LISP. The MUP-C can convert user session information from the control plane into the user plane's network protocol routing information without involving a traffic anchor entity like UPF PSA. Figure 5 describes the DMM MUP architecture with SR IPv6 (SRv6) as the underlay network protocol. The DMM MUP architecture comprises an MUP-C and multiple Provider Edge (PE) nodes. The PE nodes are the routing entities that represent the N3 and N6 interfaces. For example, PEs can be the first SRv6 node that connects to gNodeB and DN. The DMM MUP Architecture also defines four Route Types: Interwork Segment Discovery route, Direct Segment Discovery route, Type 1 Session Transformed route, and Type 2 Session Transformed route. PE uses the Interwork Segment Discovery route and Direct Segment Discovery route to advertise the N3 and N6 interface's SR reachability to the underlay SRv6 network, respectively. At the control plane, the MUP-C transforms the user session information received from the 5G Control Plane into SRv6 routing information. This information is represented by two types of Session Transformed routes, which are sent from the MUP-C to PEs. Type 2 Session Transformed route identifies the Direct Segment Discovery route that PE needs to use for uplink traffic. Meanwhile, Type 1 Session Transformed route identifies the Interwork Segment Discovery route that PE needs to use for downlink traffic. Based on this routing resolution method, a direct SRv6 user plane data path is established without traversing through an UPF. At the PE on the N3 interface side, the GTP packet's source address (gNB) and destination address (UPF) are translated to SRv6 routing segment lists from the N3 side-PE to the N6 side-PE. At the PE on the N6 interface side, the IP packet's source address (service instance) and destination address (UE) are translated to SRv6 routing segment lists in the opposite direction. After the packet reaches the destination PE, the SRv6 header is removed, and the packet is forwarded to the connecting gNB or DN of the PE.

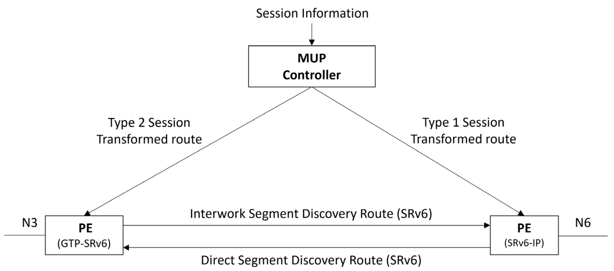
Session Information
MUP Controller
Type 2 Session Transformed route
Type 1 Session Transformed route
N3
PE (GTP-SRv6)
Interwork Segment Discovery Route (SRv6)
PE (SRv6-IP)
N6
Direct Segment Discovery Route (SRv6)

FIGURE 5. SRv6 MUP architecture.

We apply this MUP-C concept to our 5G CATS deployment solutions. This concept allows us to configure the optimal UPF-bypassing 5G user plane data path based on CATS algorithm decision using advanced underlay network protocol such as SRv6. Furthermore, it also opens the opportunity to directly configure CATS decisions at the user plane by implementing the CATS algorithm at the MUP-C. We will describe this implementation option in the next section.

III. PROPOSED 5G MOBILE USER PLANE CATS ARCHITECTURE IMPLEMENTATION OPTIONS

In this section, we discuss our 3 proposed implementation options for enabling CATS in 5G Mobile User Plane. These 3 options are called: CATS AF traffic influence, CATS MUP-C dummy-UPF, and CATS MUP-C non-UPF.

Our three proposed options share one similarity. They all use anycast address to represent the multi-instance service and MUP-C to convert the 5G control plane PDU session information into the user plane routing path. Using anycast address, a service's Fully Qualified Domain Name (FQDN) is mapped to a single IP anycast address. This address can be cached in UE during application client installation.

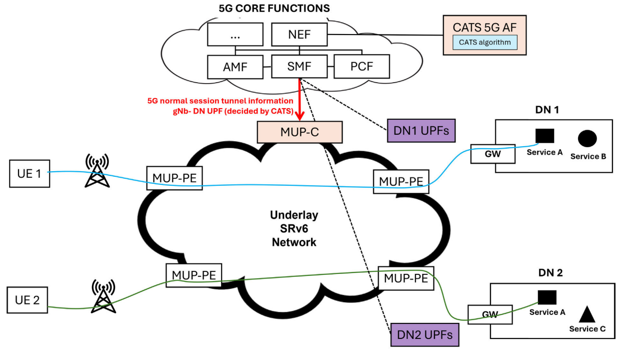

FIGURE 6. CATS AF traffic influence implementation option architecture.

UE can directly use this anycast address to request the multi-instance service. Hence, DNS resolution is not required. The 5G network is responsible for selecting and routing the anycast IP service request to an optimal service instance. This process is transparent to UE. Meanwhile, as explained in the previous section, the MUP-C concept enables direct user plane routing path configuration using SRv6, and CATS algorithm implementation option at the data plane.

The differences between the 3 proposed 5G CATS implementation options lie in where the CATS algorithm is deployed (control plane/data plane) and how the UE PDU session information is configured and provided to the MUP-C. In the following subsections, we introduce the 3 implementation options sequentially based on the increasing order of changes they make to the current 5G architecture.

A. CATS AF TRAFFIC INFLUENCE OPTION

Figure 6 illustrates the overview 5G architecture when applying the CATS AF traffic influence implementation option. In this option, the CATS algorithm is deployed at the 5G AF and provides the optimal service instance location (can be represented by a Data Network Access Identifier (DNAI)) to 5G core functions based on the AF traffic influence procedure defined by 3GPP [22]. However, 3GPP defines the AF traffic influence procedure for edge relocation use-case. In our work, we only referred to the original procedure and modified the procedure's message data for the CATS use case. Based on the CATS AF decision, SMF selects and configures the corresponding UPFs (including both UPF UL-CL and UPF PSA) of the decided service instance location. SMF then sends the configured session information to the MUP-C to convert it to the optimal SRv6 user plane routing path.

Figure 7 describes the detailed PDU session setup procedure when applying the CATS AF traffic influence implementation option.

In case of a new PDU session establishment, the CATS AF creates an AF traffic influence rule at SMF that subscribes to the PDU session Establishment event of any UE requests targeting the anycast IP address of the service. This rule should be created before the service starts to accept UE requests. When the SMF recieves any new PDU Session Establishment request from UE that targeting the configured anycast IP address, it sends a notification that includes the requested UE location and IP address information to the CATS AF. 5G network provider can select their own preferred method for representing the UE location information (e.g., UE IP address or gNB IP address, etc.). Then, the CATS algorithm running inside the CATS AF determines the optimal service instance and its DNAI based on the received UE location, computing, and networking information. The CATS AF provide this CATS decision to SMF via a new AF traffic influence rule that targeting an individual UE. This rule defines the chosen DNAI as the forwarding destination for any traffic from the requested UE (identified by the UE IP address that SMF sends to CATS AF) targeting the service anycast IP address. This rule also includes the CATS AF subscription request to SMF on any further UE path change event of the defined UE in the rule. This UE path change subscription allows the CATS AF to re-select the optimal service instance to serve the UE in case of UE moving and attaching to another gNB. Based on the new AF traffic influence rule, the SMF configures the corresponding UPFs of the chosen DNAI. Then, SMF sends the session information which includes the Tunnel Endpoint Identifier (TEID), the UE attached gNB and UPFs IP addresses to the MUP-C. The MUP-C converts the session information into the Session Transformed routes and advertises them to the 2 PEs that connect to the corresponding gNB and DN. The underlay routing path setup is completed after this step. After receiving the setup completion notification from MUP-C, the SMF finishes the remaining radio resource and UE configuration steps.

In case of PDU session modification caused by UE movement, the same procedure in Figure 7 is used. However, the new traffic influence rule targeting the individual UE created at the PDU session establishment process overrides the previous rule. Hence, the AF subscription condition at step 3 is only triggered when the targeted UE requests the anycast service at a different location (UE path change). This is the only procedure difference between the PDU establishment and modification cases.

B. CATS MUP-C DUMMY-UPF OPTION

Figure 10 illustrates the overview 5G architecture when applying the CATS MUP-C dummy-UPF option. In this option, the CATS algorithm is deployed inside the MUP-C at the 5G user plane. SMF configures a special dummy UPF to obtain the PDU session information. This dummy UPF does not have data plane routing capabilities between UE and DN. It is only capable of communicating with the SMF to create the TEID for the PDU session. The CATS-MUP-C converts the PDU session information provided by SMF to the actual optimal SRv6 routing path based on the decision output from the CATS algorithm.

Figure 8 describes the detailed PDU session setup procedure when applying the CATS MUP-C dummy-UPF option. In this deployment option, there is no difference between

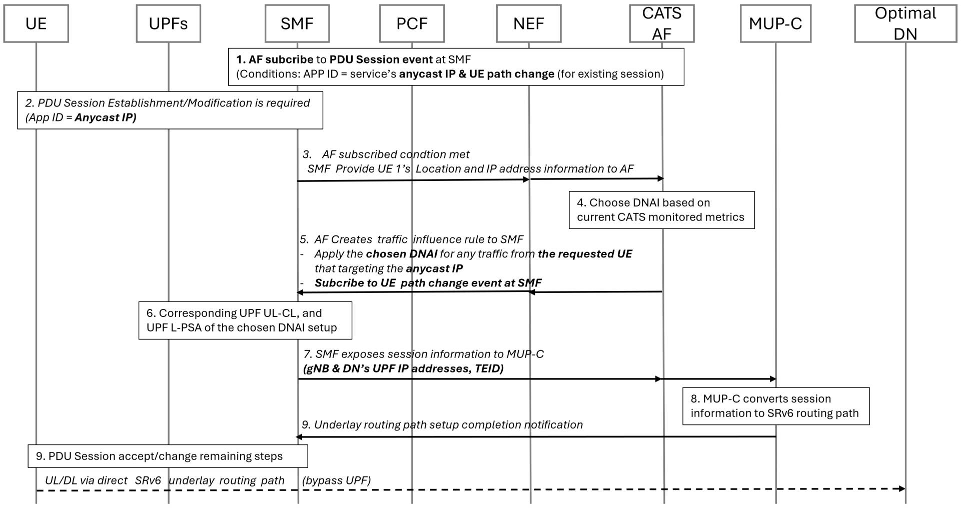
UE
UPFs
SMF
PCF
NEF
CATS AF
MUP-C
Optimal DN
1. AF subcribe to PDU Session event at SMF
(Conditions: APP ID = service's anycast IP & UE path change (for existing session)
2. PDU Session Establishment/Modification is required
(App ID = Anycast IP)
3. AF subscribed condtion met
SMF Provide UE 1's Location and IP address information to AF
4. Choose DNAI based on
current CATS monitored metrics
5. AF Creates traffic influence rule to SMF
- Apply the chosen DNAI for any traffic from the requested UE
that targeting the anycast IP
- Subcribe to UE path change event at SMF
6. Corresponding UPF UL-CL, and
UPF L-PSA of the chosen DNAI setup
7. SMF exposes session information to MUP-C
(gNB & DN's UPF IP addresses, TEID)
8. MUP-C converts session
information to SRv6 routing path
9. Underlay routing path setup completion notification
9. PDU Session accept/change remaining steps
UL/DL via direct SRv6 underlay routing path (bypass UPF)

FIGURE 7. CATS AF traffic influence implementation option PDU session setup procedure.

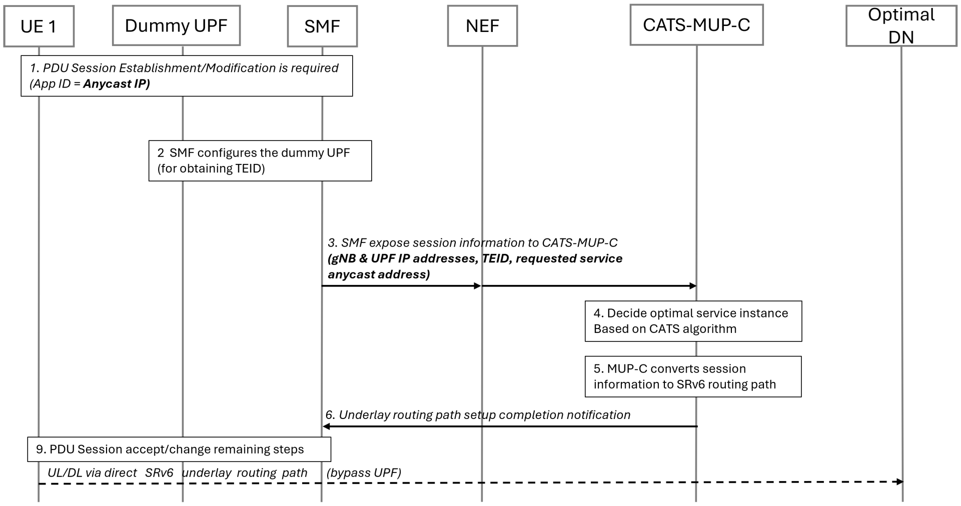
UE 1
Dummy UPF
SMF
NEF
CATS-MUP-C
Optimal DN
1. PDU Session Establishment/Modification is required
(App ID = Anycast IP)
2 SMF configures the dummy UPF
(for obtaining TEID)
3. SMF expose session information to CATS-MUP-C
(gNB & UPF IP addresses, TEID, requested service
anycast address)
4. Decide optimal service instance
Based on CATS algorithm
5. MUP-C converts session
information to SRv6 routing path
6. Underlay routing path setup completion notification
9. PDU Session accept/change remaining steps
UL/DL via direct SRv6 underlay routing path (bypass UPF)

FIGURE 8. CATS MUP-C dummy-UPF option PDU session setup procedure.

a new PDU session establishment case and an existing PDU session modification case caused by UE movement. We keep the same 5G original PDU session establishment/modification procedure until the UPF configuration step. SMF configures the dummy UPF to obtain the session's TEID. After UPF configuration, the PDU session information at SMF comprises the UE-attached gNB, dummy UPF's IP addresses, and the TEID. SMF sends this session information and the requested anycast service IP address to the CATS-MUP-C at the user plane, then waits for the CATS-MUP-C's routing path setup acknowledgment. The CATS algorithm inside the CATS-MUP-C considers the current computing and networking information related to the requested anycast service to determine the optimal service instance location. Then, the MUP-C component of the CATS-MUP-C maps the routing path between gNB and SMF's selected UPFs to the SRv6 routing path between the 2 PEs that connect to the gNB and the optimal service instance location chosen by the CATS algorithm. The CATS-MUP-C configures the corresponding Session Transformed routes to the PEs. Then, it notifies the SMF about the completion of the underlay routing path setup. The remaining radio resource and UE configuration steps complete the PDU session establishment/modification procedure.

C. CATS MUP-C NON-UPF OPTION

Figure 11 illustrates the overview 5G architecture when applying the CATS MUP-C non-UPF option. This implementation option makes the most changes to the original 5G architecture. The difference between this option and the previous CATS-MUP-C option is that the UPF configuration step is removed from the PDU session setup procedure. Because the user plane SRv6 routing path is not dependent on the UPF selected by SMF, UPF configuration is unnecessary. However, the MUP-C requires PDU session information as input. Hence, we extend the SMF with a function that

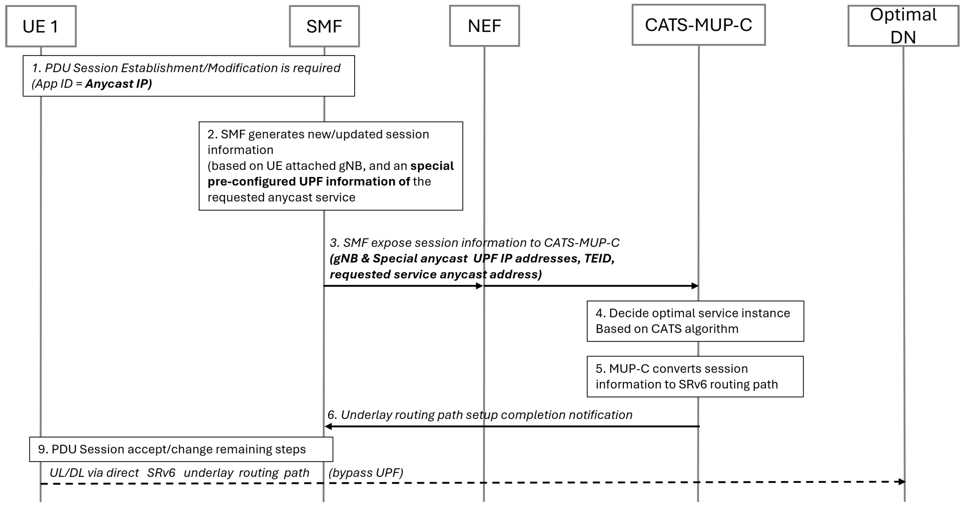
UE 1
SMF
NEF
CATS-MUP-C
Optimal DN
1. PDU Session Establishment/Modification is required (App ID = Anycast IP)
2. SMF generates new/updated session information (based on UE attached gNB, and an special pre-configured UPF information of the requested anycast service
3. SMF expose session information to CATS-MUP-C (gNB & Special anycast UPF IP addresses, TEID, requested service anycast address)
4. Decide optimal service instance Based on CATS algorithm
5. MUP-C converts session information to SRv6 routing path
6. Underlay routing path setup completion notification
9. PDU Session accept/change remaining steps
UL/DL via direct SRv6 underlay routing path (bypass UPF)

FIGURE 9. CATS MUP-C non-UPF information option PDU session setup procedure.

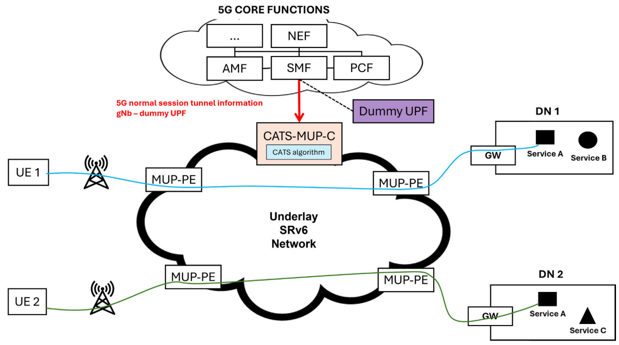

FIGURE 10. CATS MUP-C dummy-UPF option architecture.

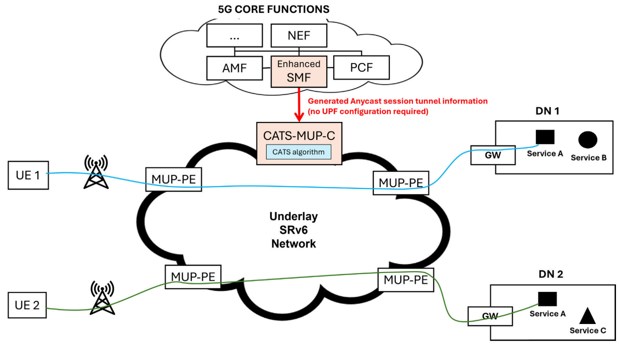

FIGURE 11. CATS MUP-C non-UPF option architecture.

can generate the PDU session information for the anycast IP address service traffic. This design allows SMF to obtain the PDU session information without performing UPF selection and configuration processes. Then, the CATS-MUP-C converts this session information to the optimal SRv6 routing path like the previous CATS-MUP-C implementation option.

Figure 9 describes the detailed PDU session setup procedure when applying the CATS MUP-C non-UPF option. We only describe the differences in steps 2 and 3 compared with the previous CATS-MUP-C dummy-UPF option. The other steps are similar. In steps 2 and 3, SMF's generated session information comprises the TEID, the UE attached gNb IP address, and a special pre-configured UPF IP address of the requested anycast IP address service. A single corresponding UPF IP address is pre-configured at the SMF for each anycast IP address service. In case of UE mobility, the SMF only needs to update the new UE attached gNB IP address. The special UPF information is unchanged. SMF then sends the updated session information to the CATS-MUP-C. By removing the UPF selection and configuration processes, the PDU session setup time in this implementation option can be reduced compared with previous options. However, it causes significant 5G architecture changes.

IV. IMPLEMENTATION

In this section, we describe our implementation setup details. We created a testbed to implement our 3 5G CATS implementation options and the 3GPP's dynamic session breakout edge discovery method.

A. CATS CONTROLLER SETUP

We define the CATS Controller (CATS-C) as the functional entity that hosts the CATS algorithm and manages the up-to-date information of the deployed services, including service discovery, computing, and networking information. Figure 12 illustrates how we implemented the CATS controller.

Regarding the CATS algorithm, this study's scope is the 5G architecture and procedure changes instead of designing an optimal algorithm. Hence, we only implemented a simple algorithm that determines the most available CPU resource service site as the optimal service instance location. We deployed a service metrics agent at each service site and used Prometheus [23] monitoring tool to collect the site's CPU resource availability.

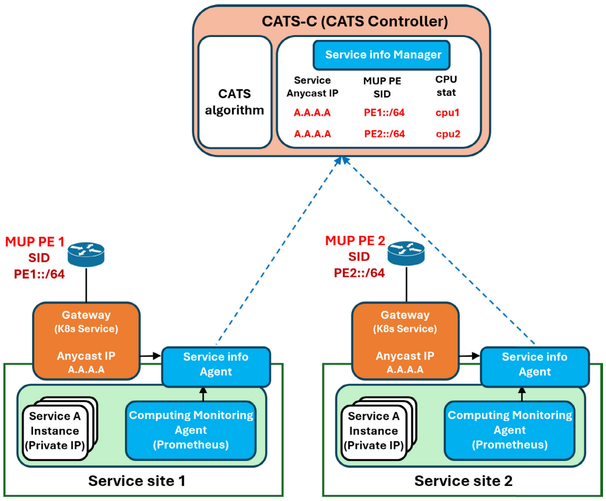

FIGURE 12. **CATS Controller implementation details.**

Regarding the service information management function, we developed a simple system that consists of an information manager at the CATS controller and an information agent at each service site. The service information manager stores the anycast IP address, the connecting MUP PE Segment ID (SID), and the current CPU status of each service. The service anycast IP address is the exposed IP of the Kubernetes Service resource that acts as a gateway for all instances (Kubernetes pods) of the same service. Because we use the anycast IP address concept, we configured the same Kubernetes service IP for the same service running at different service sites. The MUP PE SID is the IPv6 address of the PE connecting to the service site. This SID is fixed and is pre-configured at each service site. The CPU status is the service site's current CPU usage data the Prometheus agent collects. The service information agent constantly updates the information of all services running at each service site to the information manager.

Depending on design of each 5G CATS implementation option, this CATS-C component can run at the AF or at the CATS-MUP-C. In case of the 3GPP edge discovery method, this CATS-C component can run at the DNS server.

B. TESTBED SETUP

Figure 13 illustrates the overall architecture of our testbed. This general architecture shows the components of all 5G CATS implementation options that we evaluate in this study (our three proposed options and the 3GPP option). The components drawn by the black dashed line belong to different implementation options. In each option experiment, only the corresponding components are used. The unrelated components are disconnected from the testbed.

- the 3GPP edge discovery option uses the normal SMF, the MUP-C and the CATS DNS server.
- The CATS AF traffic influence option uses the normal SMF, the CATS AF, and the MUP-C.
- The CATS MUP-C dummy-UPF option uses the normal SMF and the CATS MUP-C.
- The CATS MUP-C non-UPF option uses the enhanced SMF and the CATS MUP-C

Regarding the 5G network functions, we used Free5GC [24] for deploying them. We deployed a simplified 5G network with only a single UPF anchoring each service site instead of several ULCL and PSA UPFs. This modification is reasonable in this study because the UPF configuration steps in all 5G CATS deployment options are the same. It does not affect the performance comparison results between these options. We simulated the enhanced SMF by pre-configuring the generated session information inside the default SMF.

Regarding the underlay network infrastructure, we used Mininet [25] with P4-enabled Behavioral Model version 2 (P4 BMv2) switches [26] to simulate. Our simulated network can support both GTP and SRv6 network protocols. The P4 switches connecting to gNB and service site are considered as the MUP PEs. We used the ONOS SDN controller [27] to implement the MUP-C. The ONOS SDN controller can inject the SRv6 routing policies to the P4 BMv2 MUP PE switches via P4 programming. We also developed a simple program that allows the MUP-C to receive session information from the SMF via an open restful API.

Table 1 explains how the Direct Segment Discovery and the Type 2 Session Transformed Routes are created based on our testbed environment. The MUP PE connecting to the gNB imports these two route types. Whenever this MUP PE receives the GTP packet from UE, it selects the Type 2 Session Transformed route that has the corresponding UPF address and N6 TEID with the GTP packet header. Then, it resolves the Type 2 Session Transformed route by the Direct Segment Discovery route with the same Direct Segment ID. The MUP PE SID in the corresponding Direct Segment Discovery Route identifies the target MUP PE connecting to the service site to which the packet should be forwarded.

Table 2 explains how the Interwork Segment Discovery and the Type 1 Session Transformed Routes are created based on our testbed environment. The MUP PE connecting to the service site imports these two route types. Whenever this MUP PE receives the replying packet from the service site, it selects the Type 1 Session Transformed route that has the corresponding UE address with the destination address field inside the packet header. Then, it resolves the Type 1 Session Transformed route by the Interwork Segment Discovery route with the same gNB address. The MUP PE SID in the corresponding Interwork Segment Discovery Route identifies the target MUP PE connecting to the gNB to which the packet should be forwarded.

C. EVALUATION EXPERIMENTS

We conducted two experiments to compare the performance between 4 5G CATS implementation options (3GPP dynamic edge discovery method and our three proposed options). We evaluated them regarding PDU session setup latency and routing path throughput.

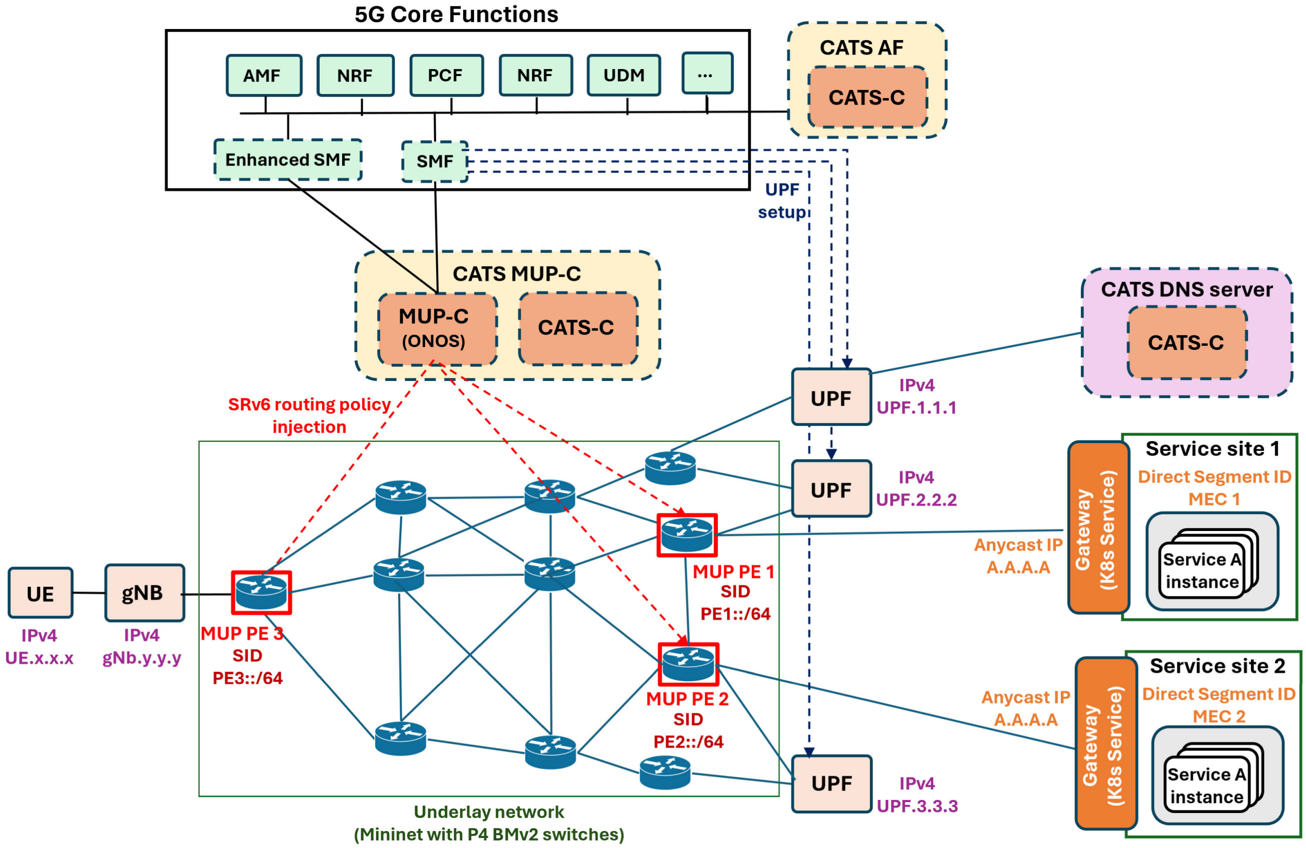

FIGURE 13. Testbed architecture for the proposed 5G CATS and 3GPP edge discovery implementation options. The deployment status of the black dashed line components depends on the chosen option.

TABLE 1. Implemented Direct Segment Discovery and Type 2 Session Transformed routes parameters.

Direct Segment Discovery Route	Type 2 Session Transformed Route
Direct Segment ID (MEC i)	Direct Segment ID (MEC i)
MUP PE SID (PE k::/64)	UPF address (UPF.j.j.j)
	N6 TEID

TABLE 2. Implemented Interwork Segment Discovery and Type 1 Session Transformed routes parameters.

Interwork Segment Discovery Route	Type 1 Session Transformed Route
gNB address (gNB.y.y.y)	gNB address (gNB.y.y.y)
MUP PE SID (PE k::/64)	UE address (UE.x.x.x)
	N3 TEID

For the PDU session setup latency comparison, the latency measurement starts when the SMF begins to process the PDU session setup request. The measurement stops when the UE can start forwarding packets to the service site via the user plane routing path. We also measured the latency of each step inside the PDU session setup procedure of each 5G CATS implementation option. Based on the design of these options, we divided their PDU session setup procedure into smaller steps, as depicted in Table 3.

Specifically, UPF configuration is the latency of selecting and configuring the UPF carried out by the SMF. DNS query/response over UPF is the latency of sending and receiving DNS query and response between UE and the DNS server. CATS service instance resolution is the latency of determining the optimal service instance location carried out by the CATS algorithm. AF traffic influence rule creation is the latency of inserting an AF traffic influence rule into 5G network. Session information generation is the latency of generating PDU session information based on gNB and pre-configured UPF information from the Enhanced-SMF. MUP-C SRv6 routing path conversion is the latency of converting the PDU session information and injecting SRv6 routing policies to the PE carried out by the MUP-C.

In each implementation option experiment, we used Kubernetes to deploy two web service instances at two servers. Both service instances can be accessed via a single anycast IP address. The UE used this pre-configured anycast IP address to request the service. We used a CPU stress tool to create a high CPU usage situation at one server so that the CATS algorithm is supposed to choose the other

TABLE 3. PDU session setup procedure step breakdown.

Steps	3GPP dynamic edge discovery	CATS AF traffic influence	CATS MUP-C dummy-UPF	CATS MUP-C non-UPF
UPF configuration	x	x	x	
Session information generation				x
DNS query/response over UPF	x			
AF traffic influence rule creation		x		
CATS service instance resolution	x	x	x	x
MUP-C SRv6 routing path conversion		x	x	x

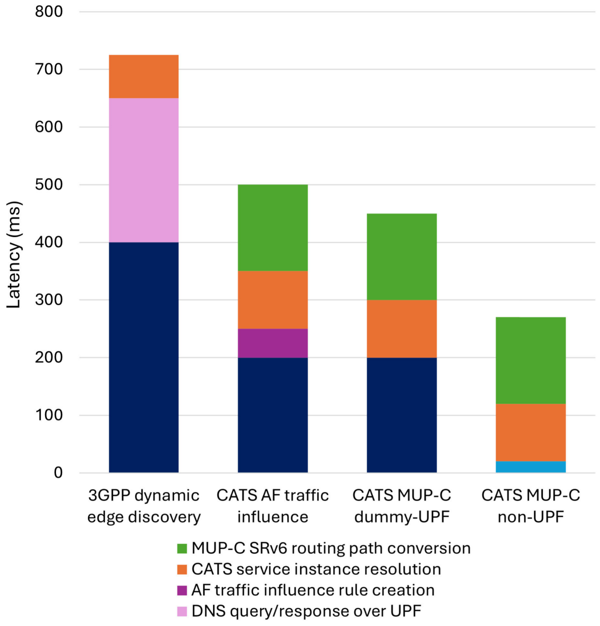

FIGURE 14. PDU Session setup latency comparison result.

server to route the request. We used timestamp logs in all 5G network functions, CATS-C and MUP-C, to measure each PDU session setup step latency. Each experiment is conducted 20 times, and the average latency over these experiment runs is recorded as the final result.

For the routing path throughput experiment, we compared the data throughput between using GTP encapsulation in 3GPP options and using SRv6 encapsulation in our proposed options because GTP and SRv6 packets have different packet sizes. Because all of the proposed options use MUP-C to replace the GTP tunnel with the SRv6 routing path, we only need to use the performance of one option for comparison. The CATS MUP-C non-UPF option was used in this experiment. We used iPerf3 to record the throughput of packets sent from UE to the service site server via the GTP tunnel and MUP-C SRv6 paths. We conducted the experiment using the smallest, middle, and largest benchmarking ethernet packet sizes (64, 512, and 1518 bytes) defined in the IETF RFC 2544 document [28].

V. RESULTS AND DISCUSSIONS

A. PDU SESSION SETUP LATENCY

Figure 14 shows the PDU session setup latency comparison between our three proposed 5G CATS implementation options and the 3GPP dynamic edge discovery option. The 3GPP option had the highest latency. The CATS AF traffic influence option's latency was approximately 30% lower than the 3GPP option. It is followed by a slightly lower latency from the CATS MUP-C dummy-option option. The CATS MUP-C non-UPF option has the lowest latency, only about half of the 3GPP method's latency.

The key reason for these latency differences is the number of times UPF needs to be configured. The 3GPP option requires configuring UPF two times. First, it has to configure UPF to set up the DNS query forwarding path between the UE and the DNS server. Then, UPF configuration is required one more time to set up the service routing path after the DNS server responds the optimal service instance location based on the CATS-C decision. The DNS query/response messages forwarding process further increases final latency. Meanwhile, the CATS AF traffic influence and the CATS MUP-C dummy-UPF option only have one UPF configuration step for the service routing path. The CATS MUP-C non-UPF option is the fastest one because it does not require the UPF configuration step. In this option, the Enhanced-SMF generates the session information based on the UE-attached gNB and the pre-configured UPF information corresponding to the requested anycast IP address.

Regarding other PDU session setup steps, all four options had a similar CATS service instance resolution time because they all used CATS-C. The CATS service instance resolution latency in our three proposed options was slightly longer because it included the UE information forwarding latency between SMF and AF or MUP-C. However, the difference is negligible because these functions are directly connected. Besides, although the 3GPP solution did not have an extra MUP-C SRv6 routing path conversion latency as our three proposed solutions, it still had the highest total PDU session setup latency. The DNS resolution procedure latency is more impactful.

B. ROUTING PATH THROUGHPUT

Figure 15 shows the routing path throughput performance comparison between the 3GPP GTP encapsulation and the MUP-C SRv6 encapsulation methods. CPU strain at the UE caused by the increasing number of packets caused lower throughput for lower packet size. This phenomenon is normal in network throughput benchmarking and does not relate to this study. The important comparison result in Figure 15 is the slightly lower throughput of the SRv6 encapsulation method

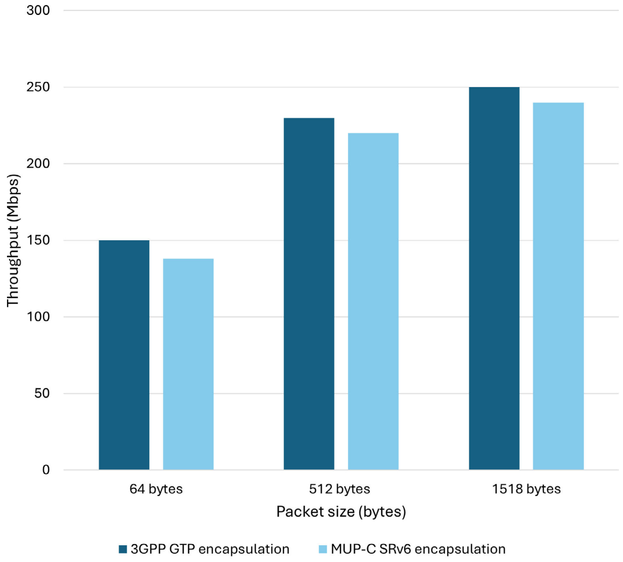

FIGURE 15. Routing path throughput comparison result.

in all packet sizes compared with the 3GPP GTP method. This phenomenon is caused by the larger size of the SRv6 packet header compared with the GTP packet header [29]. However, the difference is negligible. We can consider that SRv6 does not cause negative consequences to the routing performance in the 5G mobile user plane.

C. SUMMARY

In summary, our three proposed 5G CATS implementation solutions significantly reduced the PDU session setup latency compared with the 3GPP dynamic edge discovery method. They avoid the extra latency caused by the UPF configuration for DNS resolution and DNS query forwarding processes. The CATS MUP-C non-UPF option can further reduce the latency by removing the UPF configuration process for DN. Besides, the SRv6 encapsulation used by the MUP-C in our solutions does not cause significant packet routing throughput overhead. Besides, our proposed solutions can configure shorter routing paths than the 3GPP method because of the MUP-C bypassing UPF routing path configuration feature. We do not include this benefit evaluation in this study because it was already proved in the original MUP-C concept demonstration [20].

VI. CONCLUSION

In conclusion, this study proposed 3 5G architecture implementation options to support CATS: CATS AF traffic influence, CATS MUP-C dummy-UPF, and CATS MUP-C non-UPF. We compared the performance of our solutions with the 3GPP dynamic edge discovery procedure, which can be used to support CATS. Our proposed 5G CATS implementation options significantly reduced the PDU session setup latency by utilizing the anycast IP address service representation and the MUP-C concept. In the 3GPP method, the CATS algorithm runs on the DNS server. The DNS server's UPF configuration and DNS query forwarding steps cost extra latency. Meanwhile, our solutions do not require the DNS resolution process because UE requests the service by a single IP address. The MUP-C directly configures the routing path toward an optimal service instance location based on the information from the CATS algorithm running inside the architecture. The CATS MUP-C non-UPF option further reduced the PDU session setup latency by introducing the PDU session information generation concept without any UPF configuration required. For future works, we plan to implement the 5G CATS architecture to support the service function chaining use-case.

REFERENCES

[1] K. Yao et al. (Jan. 2, 2024). *Computing-Aware Traffic Steering (CATS) Problem Statement, Use Cases, and Requirements*. Internet Eng. Task Force CATS Working Group. [Online]. Available: https://datatracker.ietf.org/doc/draft-ietf-cats-usecases-requirements/

[2] *5G System Enhancements for Edge Computing; Stage 2*, Standard TS 23.548, 3GPP Technical Specification (TS) Release 18, version 18.5.0, 2024.

[3] S. Matsushima et al. (Mar. 3, 2024). *Mobile User Plane Architecture for Distributed Mobility Management*. Internet Engineering Task Force DMM Working Group. [Online]. Available: https://datatracker.ietf.org/doc/draft-mhkk-dmm-mup-architecture/00/

[4] C. Li et al. (Apr. 30, 2024). *A Framework for Computing-Aware Traffic Steering (CATS)*. Internet Engineering Task Force CATS Working Group. [Online]. Available: https://datatracker.ietf.org/doc/draft-ietf-cats-framework/

[5] A. Lacava, M. Polese, R. Sivaraj, R. Soundrarajan, B. S. Bhati, T. Singh, T. Zugno, F. Cuomo, and T. Melodia, "Programmable and customized intelligence for traffic steering in 5G networks using open RAN architectures," *IEEE Trans. Mobile Comput.*, vol. 23, no. 4, pp. 2882–2897, Apr. 2023.

[6] V.-D. Nguyen, T. X. Vu, N. T. Nguyen, D. C. Nguyen, M. Juntti, N. C. Luong, D. T. Hoang, D. N. Nguyen, and S. Chatzinotas, "Network-aided intelligent traffic steering in 6G O-RAN: A multi-layer optimization framework," *IEEE J. Sel. Areas Commun.*, vol. 42, no. 2, pp. 389–405, Feb. 2024.

[7] F. Kavehmadavani, V.-D. Nguyen, T. X. Vu, and S. Chatzinotas, "Intelligent traffic steering in beyond 5G open RAN based on LSTM traffic prediction," *IEEE Trans. Wireless Commun.*, vol. 22, no. 11, pp. 7727–7742, Mar. 2023.

[8] A. Thantharate, V. Walunj, R. Abhishek, and R. Paropkari, "Balanced5G–fair traffic steering technique using data-driven learning in beyond 5G systems," in *Proc. 8th Int. Conf. Adv. Comput. Commun. Syst. (ICACCS)*, vol. 1, Coimbatore, India, Mar. 2022, pp. 1–7.

[9] H. Zhang, H. Zhou, M. Elsayed, M. Bavand, R. Gaigalas, Y. Ozcan, and M. Erol-Kantarci, "On-device intelligence for 5G RAN: Knowledge transfer and federated learning enabled UE-centric traffic steering," *IEEE Trans. Cognit. Commun. Netw.*, vol. 10, no. 2, pp. 689–705, Apr. 2024.

[10] I. Chatzistefanidis, N. Makris, V. Passas, and T. Korakis, "ML-based traffic steering for heterogeneous ultra-dense beyond-5G networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Glasgow, U.K., Mar. 2023, pp. 1–6.

[11] I. Chatzistefanidis, N. Makris, V. Passas, and T. Korakis, "Which ML model to choose? Experimental evaluation for a beyond-5G traffic steering case," in *Proc. ICC IEEE Int. Conf. Commun.*, Rome, Italy, May 2023, pp. 5185–5190.

[12] X. Ba, L. Jin, Z. Li, J. Du, and S. Li, "Multiservice-based traffic scheduling for 5G access traffic steering, switching and splitting," *Sensors*, vol. 22, no. 9, p. 3285, Apr. 2022.

[13] H. Erdol, X. Wang, P. Li, J. D. Thomas, R. Piechocki, G. Oikonomou, R. Inacio, A. Ahmad, K. Briggs, and S. Kapoor, "Federated meta-learning for traffic steering in O-RAN," in *Proc. IEEE 96th Veh. Technol. Conf. (VTC-Fall)*, London, U.K., Sep. 2022, pp. 1–7.

[14] J. Zeng, H. Wang, and W. Luo, "Self-optimizing traffic steering for 5G mmWave heterogeneous networks," *Sensors*, vol. 22, no. 19, p. 7112, Sep. 2022.

[15] P. V. Wadatkar, R. G. Garroppo, G. Nencioni, and M. Volpi, "Joint multi-objective MEH selection and traffic path computation in 5G-MEC systems," *Comput. Netw.*, vol. 240, Feb. 2024, Art. no. 110168.
[16] S. D. A. Shah, M. A. Gregory, S. Li, R. d. R. Fontes, and L. Hou, "SDN-based service mobility management in MEC-enabled 5G and beyond vehicular networks," *IEEE Internet Things J.*, vol. 9, no. 15, pp. 13425–13442, Aug. 2022.
[17] M. R. Anwar, S. Wang, M. F. Akram, S. Raza, and S. Mahmood, "5G-enabled MEC: A distributed traffic steering for seamless service migration of Internet of Vehicles," *IEEE Internet Things J.*, vol. 9, no. 1, pp. 648–661, Jan. 2022.
[18] S. Matsushima et al., *Segment Routing Over IPv6 for the Mobile User Plane*, document RFC 9433, Internet Eng. Task Force, Jul. 2023. [Online]. Available: https://datatracker.ietf.org/doc/html/rfc9433
[19] Z. Zhang et al. (Feb. 23, 2024). *Mobile User Plane Evolution*. Internet Engineering Task Force DMM Working Group. [Online]. Available: https://datatracker.ietf.org/doc/draft-zzhang-dmm-mup-evolution/
[20] S. Matsushima, "SRv6MUP A Mobile User Plane Network Evolution with SRv6," Presented at MPLS SD AI Net World Congress, 2022. [Online]. Available: https://www.segment-routing.net/conferences/MPLS-WC-2022-Satoru-Matsushima/
[21] K. Samdanis and T. Taleb, "The road beyond 5G: A vision and insight of the key technologies," *IEEE Netw.*, vol. 34, no. 2, pp. 135–141, Mar. 2020.
[22] *Procedures for the 5G System (5GS)*, Standard TS 23.502, 3GPP Technical Specification (TS) Release 18, version 18.5.0, 2024.
[23] *Prometheus—Monitoring System & Time Series Database*. [Online]. Available: https://prometheus.io/
[24] *Free5GC*. [Online]. Available: https://free5gc.org/
[25] *Mininet—An Instant Virtual Network on Your Laptop (or Other PC)*. [Online]. Available: https://mininet.org/
[26] *Behavioral Model (BMV2)*. [Online]. Available: https://github.com/p4lang/behavioral-model
[27] *Open Network Operating System (ONOS) SDN Controller for SDN/NFV Solutions*. [Online]. Available: https://opennetworking.org/onos/
[28] S. Bradner and J. McQuaid, *Benchmarking Methodology for Network Interconnect Devices*, document RFC 2544, Mar. 1999. [Online]. Available: https://www.rfc-editor.org/rfc/rfc2544.html
[29] C. Lee, K. Ebisawa, H. Kuwata, M. Kohno, and S. Matsushima, "Performance evaluation of GTP-U and SRv6 stateless translation," in *Proc. 15th Int. Conf. Netw. Service Manage. (CNSM)*, Halifax, NS, Canada, Oct. 2019, pp. 1–6.

MINH-NGOC TRAN received the B.E. degree from Hanoi University of Science and Technology, in 2018, and the M.S. degree from Soongsil University, in 2020, where he is currently pursuing the Ph.D. degree in information and communication convergence. He works on research projects related to cloud/edge computing, serverless computing, and 5G/6G network infrastructure.

VAN-BINH DUONG received the B.Sc. degree in data science from the University of Information Technology, VNU-HCM, in 2022. He is currently pursuing the M.Sc. degree with Soongsil University, Seoul, South Korea. He works on research projects related to cloud computing and 5G/6G network infrastructure.

YOUNGHAN KIM (Member, IEEE) received the B.S. degree from Seoul National University, South Korea, and the M.Sc. and Ph.D. degrees in electrical engineering from Korea Advanced Institute of Science and Technology (KAIST). He was the President of Korea Information and Communications Society (KICS), in 2021. He is currently a Full Professor with the Department of Electronic Engineering, Soongsil University. His current research interests include cloud computing, 5G networking, and next-generation networks.