IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 7 August 2024, accepted 29 August 2024, date of publication 4 September 2024, date of current version 13 September 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3454251

RESEARCH ARTICLE

Front Haul Multi-Homing for Cell-Free Massive MIMO: A Low-Latency Approach With Distributed and Hierarchical CPU Deployment

AKIO IKAMI, AMR AMRALLAH, (Member, IEEE), YU TSUKAMOTO, TAKAHIDE MURAKAMI, HIROYUKI SHINBO, AND YOSHIAKI AMANO

Wireless Technology Division, KDDI Research, Inc., Saitama 356-8502, Japan

Corresponding author: Akio Ikami (ak-ikami@kddi.com)

This work was supported by the National Institute of Information and Communications Technology (NICT), Japan, under Grant JPJ012368C00401.

ABSTRACT In anticipation of the 6G system, it is crucial to address the challenges related to achieving high throughput and low latency over a wide area. This is particularly important for applications such as 4K 360-degree streaming, tactile communication, and autonomous driving. Cell-free massive MIMO (CF-mMIMO) has been proposed as a solution to enhance radio coverage by eliminating cell edges, where a central processing unit (CPU) orchestrates signal processing across distributed access points (APs). This paper explores the deployment of CF-mMIMO within a realistic hierarchical and distributed site architecture to meet the stringent latency requirements of 6G through mobile edge computing (MEC). However, the deployment of CF-mMIMO and utilization of MEC within a distributed and hierarchical architecture necessitates the placement of CPUs at edge sites. This arrangement may lead to radio quality degradation at the edges of distributed sites due to inter-site interference and AP selection limitations. To address this problem, we propose a multi-homing AP connection that allows APs to connect to multiple edge sites, thereby enabling signal processing across sites. Furthermore, we conduct an analysis of the fronthaul (FH) costs associated with multi-homing connections and propose a multi-homing configuration and multi-homing AP selection method that improves radio quality without incurring excessive FH costs. The proposed heuristic algorithm is introduced to optimize the AP connections for multi-homing, giving priority to APs near site edges to strike a balance between cost and performance. Simulation results demonstrate that our proposed method attains near-optimal throughput while preserving cost efficiency.

INDEX TERMS CF-mMIMO, virtualized RAN, multi-homing.

I. INTRODUCTION

The 6G system, slated for implementation around 2030, is anticipated to fulfill a range of requirements, including high throughput and low latency, regardless of the user equipment (UE) location. Anticipated applications for the 6G system include 4K 360-degree streaming, tactile communication, and autonomous driving [1]. To meet these demands, it is imperative to secure high-quality radio coverage for UEs distributed across a wide area. Cell-free massive multiple input and multiple output (CF-mMIMO) has been proposed as a solution to eliminate the cell edge effect, a region where radio quality typically degrades in conventional cellular systems [2]. In CF-mMIMO, a central processing unit (CPU) performs multi-user MIMO signal processing for all users, processing radio signals transmitted and received from multiple distributed access points (APs).

To implement CF-mMIMO over a wide area, AP clustering [3] and distributed site architectures have been proposed [4], [5], [6]. AP clustering mitigates the computational complexity of signal processing by serving UEs through a select number of APs, known as the AP cluster. By selecting

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

APs with high signal power for each UE, radio quality can be maintained. In a distributed site architecture, CPUs are dispersed at each edge site, serving as a physical base for aggregating the radio signals of APs in a divided area. To utilize distributed APs in each edge site, we proposed a hierarchical distributed site architecture [5] as a part of our user-centric radio access network approach. This architecture incorporates a central site designed to aggregate the radio signals from edge sites. Virtualized CPUs (vCPUs) are strategically positioned on general-purpose servers, either in the central site or edge sites, contingent upon service requirements and network conditions, such as the transmission load of backhaul links. Deploying vCPUs at edge sites is crucial for diminishing the transmission load between edge sites and the central site, as only the user data processed by the vCPUs are transmitted. This strategy aligns with the low latency requirements set forth by 6G, which can be effectively addressed through mobile edge computing (MEC) [7]. MEC necessitates the colocation of application servers with vCPUs at the edge sites to process data in closer proximity to end-users, thereby substantially reducing response times. In this paper, we explore the deployment of vCPUs at edge sites with the dual objective of ensuring uniform and stable radio quality over a wide area and meeting the low latency demanded by 6G applications. By integrating vCPUs into the MEC framework, we leverage the physical proximity to application servers, thereby minimizing data processing and communication delays, which ultimately enhances the user experience [1].

The degradation of radio quality for users located near the boundaries of distributed edge sites poses a substantial challenge. This problem stems from two primary factors. The first challenge is inter-site interference, as vCPUs at different sites cannot cooperate in signal processing. The second challenge involves the limited number of APs for the AP cluster that can be selected among those connected to the edge site. vCPUs cannot select APs with high signal power for the AP cluster if they are connected to another edge site. This results in degraded radio quality for UEs located at the border of the edge site. To address this issue, a method for sharing radio signals between edge sites was proposed [4]. In this method, the radio signal of the AP, connected to an edge site different from where the vCPU is placed, is received via the backhaul. However, the aggregation of radio signals via the backhaul could result in longer transmission delays, which may risk failing to meet the low latency requirements stipulated by MEC.

To mitigate the degradation of radio quality near the boundaries of sites when deploying vCPUs at edge sites, we consider applying a multi-homing technique [8], [9] to the fronthaul (FH) between the AP and edge sites. In this approach, an AP connects to multiple edge sites via the FH and transmits duplicated radio signals received from UEs. Since the radio signal of the multi-homing AP can be received by vCPUs deployed at multiple edge sites, it can ensure high radio quality by incorporating them into the AP cluster and executing signal processing at an edge site different from where the vCPU is placed. However, if multi-homing is implemented at numerous links between APs and edge sites to ensure high radio quality, additional FH installation becomes necessary, leading to elevated FH costs. To strike a balance between suppressing increased FH cost and improving radio quality, we propose an AP connection method for multi-homing. We aim to utilize the minimum number of total FH links to mitigate the increase in FH cost. In the proposed heuristic algorithm, we prioritize establishing connections between APs that are close to site edges, which are likely to benefit from multi-homing, while concurrently minimizing the number of FH installations. By strategically deploying FH for multi-homing purposes among these APs, we can duplicate only the necessary and sufficient radio signals for the users. This allows us to achieve an optimal balance between cost and radio quality. The contributions of this paper can be summarized as follows.

- We address the challenge of radio quality degradation near the boundaries of distributed edge sites by proposing a multi-homing technique to the FH between the AP and edge sites for sharing radio signals between edge sites for users using MEC.
- We propose a heuristic algorithm for an AP connection method for multi-homing that aims to minimize the number of total FH links required, balancing FH cost and radio quality improvement.
- We provide simulation results demonstrating our proposed method's effectiveness in balancing near-ideal throughput values and FH costs.

The rest of this paper is organized as follows. Section II outlines the system model and discusses the challenges associated with achieving both low latency and stable radio quality. In Section III, we elaborate on the proposed concept of multi-homing AP connection and describe in detail the design algorithm for the suggested multi-homing approach. Section IV delves into our simulation setup and results, including conditions that span a 1 km^2 area divided into nine sites, with APs and UEs randomly distributed within this space. The simulation results, which demonstrate our proposed method's ability to approach the ideal throughput values while reducing costs, are thoroughly explored in this section. Finally, Section V provides the conclusion and final remarks on this paper.

II. SYSTEM MODEL

A. NOTATIONS

Boldface uppercase and lowercase letters denote matrices and vectors, respectively. The superscript $(\cdot)^H$ denote the conjugate-transpose. The estimated operator is denoted by $\{\hat{\cdot}\}$ and the cardinality of the set $\mathcal{A}$ is denoted by $|\mathcal{A}|$. The $L \times L$ identity matrix is $\mathbf{I}_L$ and the Euclidean norm is denoted by $\|\cdot\|$. Table 1 summarizes the used symbols through this manuscript.

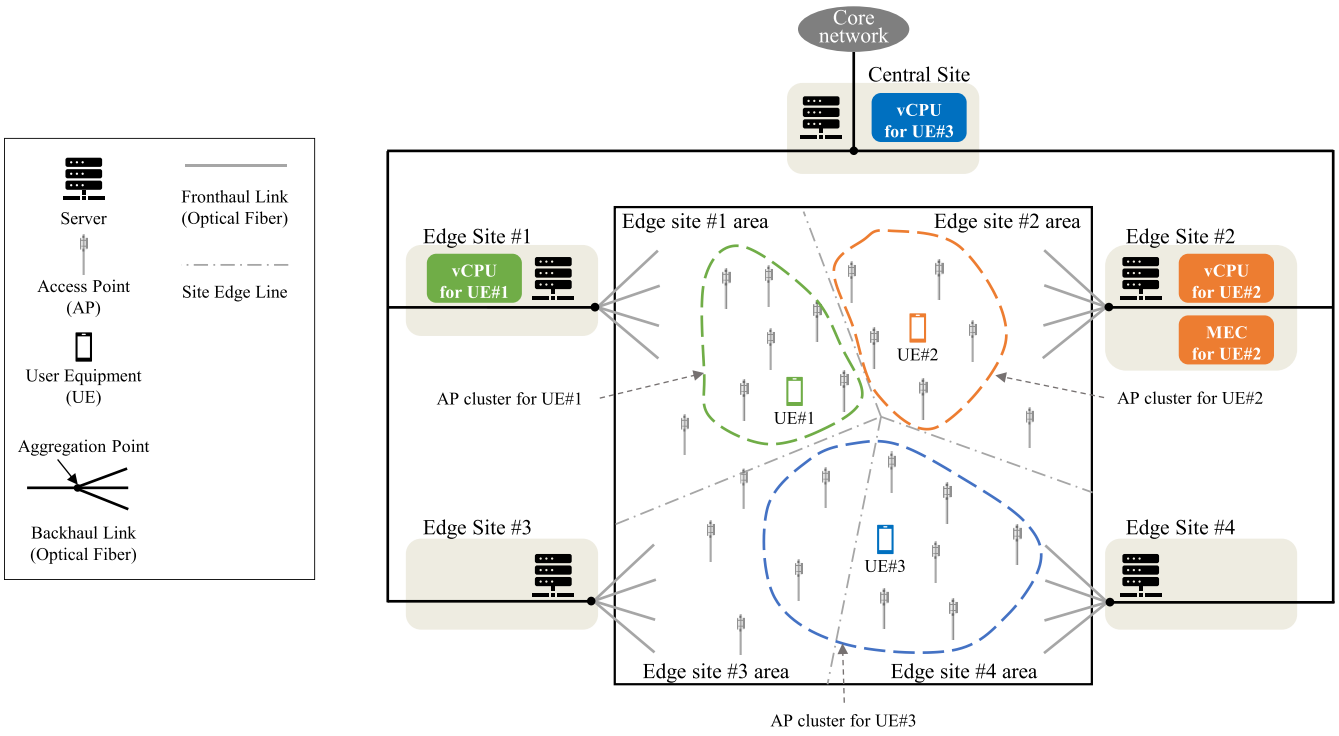


FIGURE 1. Distributed and hierarchical CPU deployment system structure for CF-mMIMO with low latency.

TABLE 1. Notations of system structures.

Symbol	Definition
L	Total number of APs
l	Index of AP
K	Total number of UEs
k	Index of UE
J	Total number of edge sites
j	Index of edge site
$x_{AP,l}$	Location of AP l
$x_{site,j}$	Signal aggregation points of edge site j
Q_l	Edge site connected to AP l
TP_k	Throughput of the uplink for UE k
$SINR_k$	SINR for UE k
W_{RF}	Wireless link bandwidth
$\boldsymbol{D}_k$	L-dimensional square matrix which indicates the AP cluster.
p_k	Power of the uplink signal
$\boldsymbol{v}_k$	Uplink combining vector for UE k
$\widehat{\boldsymbol{h}}_i$	Estimated channel coefficient of UE i
$\boldsymbol{C}_i$	Matrix of the channel estimation error for UE i
σ^2	Power of thermal noise
$\boldsymbol{I}_L$	L-dimensional identity matrix.
$\mathcal{P}_k$	Set of UEs where the AP cluster for the UE is formed with at least one AP as used in the AP cluster for UE k

TABLE 2. Notations of proposed method.

Symbol	Definition
$\boldsymbol{Q}_{\mathrm{MH},l}$	Edge site to which AP l is connected through multi-homing
x_{th}	Threshold of the distance between APs
$x_{limit-MH}$	Maximum number of APs per multi-homing
P_l	Multi-homing pairs for AP l
$AP_{dist,l}$	Distance between AP l and all APs
$targetAP_{dist,l}$	A collection of the closest ones only based on the $AP_{dist,l}$ distances
$MH_{\mathrm{AP},l}$	The group of APs that perform multi-homing with AP l

B. SYSTEM STRUCTURE

The system structure of the distributed and hierarchical site architecture assumed in this paper is depicted in Fig. 1. A double star topology, typical in an optical access network [10], [11], is adopted as the physical topology. K UEs and L APs are scattered across the communication area. For simplicity, it is assumed that each UE and AP is equipped with a single antenna. The communication area is divided into J areas using Voronoi regions, with edge sites deployed in each area. Edge sites serve as data centers, with general-purpose servers hosting vCPUs deployed at each site. Additionally, at Edge site #2, MEC is deployed. Since UE#2 utilizes MEC for processing, low-latency communication can be anticipated [7]. This paper assumes that each edge site has sufficient physical computation resources for signal processing. $x_{\mathrm{AP},l}$ represents the location of AP l, and $x_{\mathrm{site},j}$ denotes the location of the signal aggregation point of edge site j.

It is assumed that the aggregation points of each edge site are randomly deployed within the respective communication area. FHs, with connection components linked between AP l and edge site Q_l, are installed along the x-axis and y-axis [12]. The vCPUs of each UE are placed at edge sites. To reduce the computational complexity of signal processing, vCPUs of each UE select APs to be used as an AP cluster [3] from among the APs connected to the edge site where they are located.

The throughput of the uplink for UE k is denoted by TP_k, and can be defined as:

$$TP_k = W_{\mathrm{RF}} \log_2 (1 + \mathrm{SINR}_k), \tag{1}$$

where W_{RF} is the wireless link bandwidth. SINR_k is the signal-to-interference-plus-noise ratio (SINR) of UE k, which is defined as:

$$\mathrm{SINR}_k = \frac{p_k \left| \boldsymbol{v}_k^H \boldsymbol{D}_k \hat{\boldsymbol{h}}_k \right|^2}{\sum_{i=1, i\neq k}^{K} p_i \left| \boldsymbol{v}_k^H \boldsymbol{D}_k \hat{\boldsymbol{h}}_i \right|^2 + \boldsymbol{v}_k^H \boldsymbol{Z}_k \boldsymbol{v}_k}, \tag{2}$$

where p_k is the power of the uplink signal and $\boldsymbol{v}_k$ is the uplink combining vector for UE k. $\boldsymbol{D}_k$ is the L-dimensional square matrix which indicates the AP cluster. $\hat{\boldsymbol{h}}_i$ is the estimated channel coefficient of UE i. The channel coefficients are estimated with the minimum mean square error (MMSE) based on the pilot assignment method [6]. Where $\boldsymbol{Z}_k = \boldsymbol{D}_k \left(\sum_{i=1}^{K} p_i \boldsymbol{C}_i + \sigma^2 \boldsymbol{I}_L \right) \boldsymbol{D}_k$, $\boldsymbol{C}_i$ is the matrix of the channel estimation error for UE i, which is obtained from the difference between the spatial channel correlation matrix estimated with the MMSE and a real one. σ^2 is the power of thermal noise, and $\boldsymbol{I}_L$ is the L-dimensional identity matrix.

The AP index, which belongs to an AP cluster, $\boldsymbol{D}_k$, is defined as:

$$\boldsymbol{D}_k = \begin{bmatrix} D_{k1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & D_{kL} \end{bmatrix}, \tag{3}$$

where D_{kl} is defined as:

$$D_{kl} = \begin{cases} 1 & \text{if the AP } l \text{ serves UE } k, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Here, $\boldsymbol{D}_k$ is selected from the set of APs that are connected to the same edge site with the vCPU of UE k. We define selectable APs of the AP cluster from UE k as $\mathcal{L}_{U_k} = \{l | \mathcal{Q}_l = U_k\}$, where U_k is the edge site index in which the vCPU of UE k is placed.

In this paper, it is assumed that P-MMSE [6] is used for uplink signal processing for UE k. $\boldsymbol{v}_k$ is defined as:

$$\boldsymbol{v}_k = p_k \left(\sum\nolimits_{i \in \mathcal{P}_k} p_i \boldsymbol{D}_k \hat{\boldsymbol{h}}_i \hat{\boldsymbol{h}}_i^H \boldsymbol{D}_k + \boldsymbol{Z}_k \right)^{\dagger} \boldsymbol{D}_k \boldsymbol{h}_k, \tag{5}$$

where $\mathcal{P}_k$ is the set of UEs where the AP cluster for the UE is formed with at least one AP as used in the AP cluster for UE k expressed as $\mathcal{P}_k = \{i : \boldsymbol{D}_k \boldsymbol{D}_i \neq \boldsymbol{O}_L\}$ where $\boldsymbol{O}_L$ is the L-dimensional zero matrix.

C. RELATED WORKS

CF-mMIMO is emerging as a key enabler for 6G networks, offering high throughput and low latency [13]. This user-centric technology surpasses cell-edge limitations by forming clusters of transmitters around each user to ensure consistent coverage [14]. However, the deployment of CF-mMIMO presents challenges, including the management of FH capacity and accurate channel state information (CSI) estimation. FH links must be capable of sustaining the data rates required for user-centric operations [15]. Additionally, CSI is a critical parameter for beamforming and interference management. Hence, pilot assignment policies were developed to improve CSI accuracy [16].

The formation of serving clusters is a key element of user-centric CF-mMIMO, which dynamically associates each user with a subset of APs, providing the strongest signal. This approach significantly deviates from traditional cell-free massive MIMO systems where, theoretically, all transmitters can serve any user [14], [17], [18]. This dynamic clustering process is optimized based on various network performance metrics, such as scheduling or power allocation algorithms, thereby enhancing signal quality and reducing complexity [19], [20], [21], [22]. By leveraging user location information, the system intelligently forms clusters, minimizing the reliance on full CSI [23].

To tackle the complex joint optimization of AP clustering and beamforming design, deep reinforcement learning models like deep deterministic policy gradient (DDPG) and deep double Q-network (DDQN) have been introduced. These models offer improved performance while reducing computational demands [20]. Furthermore, methods using multi-agent reinforcement learning are being considered to address power consumption and scalability issues in CF-mMIMO networks. An Access Point Activation strategy based on Multi-Agent Deep Reinforcement Learning (MADRL) allows each AP to autonomously switch on/off in response to dynamic user demands as independent agents [24]. This distributed DRL method utilizes multiple learners classified based on the Reference Signal Received Power among APs surrounding a UE for each radio environment, aimed at improving the efficiency of AP clustering in large-scale CF-mMIMO [25]. Such technologies could enhance the efficiency and sustainability of cell-free massive MIMO networks, providing effective solutions to the challenges of power consumption and scalability. However, the study may overlook the computational complexity and scalability challenges in dynamic environments. Further exploration is needed on the system's performance with an increasing number of UEs and APs.

These advances in AP clustering are pivotal in enhancing the scalability and adaptability of user-centric CF-mMIMO systems, ensuring efficient resource utilization and superior network performance [4], [19], [20], [23].

Large-scale CF-mMIMO systems are set to fulfill the high throughput demands and low latency requirements of 6G networks, leveraging a distributed network of APs to overcome traditional cellular limitations [26], [27]. These

systems prioritize interference suppression, a critical aspect when CPUs are distributed without the ability to cooperate directly [26]. Distributed precoding through over-the-air signaling has been proposed to eliminate the need for backhaul CSI exchange [28], thereby enhancing scalability via dynamic clustering of APs around users [29], and power control strategies tailored to the channel conditions of individual UEs [30]. Multi-homing techniques further ensure radio quality at the peripheries of distributed sites by enabling APs to connect to multiple edge sites [31].

To maintain scalability with minimal performance degradation, a framework has been proposed whereby multiple CPUs service distinct AP clusters [31]. This approach is supported by AP clustering, which reduces computational complexity by selecting a subset of APs with high signal power to serve UEs, ensuring consistent radio quality [3]. Distributed site architectures have been proposed to facilitate CF-mMIMO deployment over wide areas, with CPUs placed across edge sites to aggregate radio signals from APs [4], [5], [6]. A hierarchical distributed site architecture has been proposed, wherein vCPUs are hosted on general-purpose servers at either central or edge sites, based on service requirements and network status [5].

However, deploying vCPUs at edge sites introduces challenges, such as inter-site interference and a limited selection of APs for AP clusters [26]. To address these challenges, methods for sharing radio signals between edge sites have been explored, enabling the reception of signals from APs connected to different edge sites [4]. While these methods aim to improve radio quality for UEs located at the borders of edge sites, it is important to consider the potential increase in transmission load between edge sites.

Resource allocation in CF-mMIMO, including power and bandwidth, is meticulously managed to accommodate the diverse needs of users. Scalability becomes a critical concern as the network expands [32]. Emerging technologies like millimeter-wave communication and software-defined networks (SDNs) are being explored for their potential to manage and support the complex operations of CF-mMIMO networks. Specifically, SDNs provide a flexible and programmable infrastructure capable of adapting to dynamic requirements of a user-centric cell-free network [33], [34].

The integration of these strategies forms the foundation for deploying scalable CF-mMIMO systems. It emphasizes user-centric coverage, dynamic clustering, advanced power control, and the incorporation of emerging technologies, all aimed at ensuring consistent service quality and network scalability.

As a new study, an approach to the challenge of massive MIMO channel estimation in mmWave communications utilizing lens antenna arrays for 5G and subsequent wireless communication systems has been proposed [35]. This research introduces two methods: a compressed sensing scheme based on convex optimization and an estimation algorithm based on the greedy method, which enable rapid reconstruction. Compared to traditional methods such as

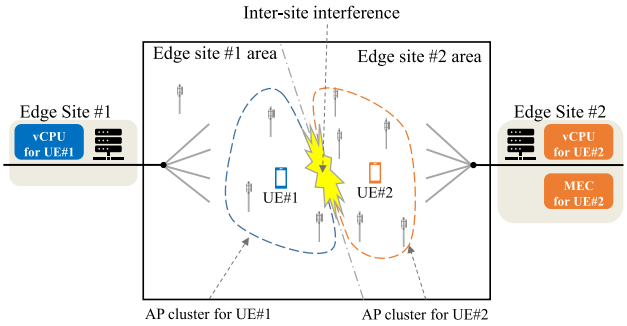


FIGURE 2. Problem of edge site vCPU placement.

support detection, orthogonal matching pursuit, and sparsity mask detection, these methods demonstrate superior performance in terms of channel estimation accuracy, recovery quality, and convergence speed. The use of lens antenna arrays highlights the importance of accurate channel estimation for beamforming, precoding, spatial multiplexing, and interference management in massive MIMO systems, contributing to system performance optimization, interference reduction, and the achievement of high data rates through reliable CSI.

Recent studies on CF-mMIMO systems based on the open radio access networks (O-RAN) architecture focus on solving scalable pilot allocation (PA) problems, energy-efficient deployment, and developing user-centric clustering and handover strategies. Specifically, a low-complexity PA scheme using MADRL has been proposed, allowing multiple learning agents to perform distributed learning on the O-RAN communication architecture to suppress pilot contamination [36]. Additionally, an approach that minimizes end-to-end power consumption by jointly allocating radio, optical fronthaul, and virtualized cloud processing resources is being explored [37]. Furthermore, by realistically modeling the temporal evolution of channels in cell-free massive MIMO and developing fixed and opportunistic clustering strategies, the implementation flexibility and interoperability of the system are enhanced [38]. These studies provide effective solutions to the challenges of power consumption and scalability in CF-mMIMO systems, representing significant advancements towards the realization of next-generation wireless networks.

D. PROBLEM STATEMENT

The AP selection process faces an inherent constraint, as the AP cluster considered for UE k is strictly confined to $\mathcal{L}_{U_k}$, which only encompasses APs within the radio signal transmission range. This constraint becomes particularly challenging for UEs located at the periphery of an edge site's coverage area. As depicted in Fig. 2, these UEs may detect APs outside of their designated set $\mathcal{L}_{U_k}$, which offer superior signal strength but are associated with a different edge site $Q_l \neq U_k$. The inability to leverage these high-power APs

in the selection process leads to a suboptimal configuration for the AP cluster, which subsequently leads to a decline in SINR.

This decline is attributed to the reduction of the signal power term in the numerator of the SINR equation (2), which is a critical factor in determining the overall quality of the wireless connection. The current architecture's limitations highlight the need for a more flexible and intelligent AP selection strategy, one that can dynamically adapt to the varying signal landscapes experienced by UEs, particularly those at the edge of a site's domain. Such a strategy would not only address the issue of maintaining high SINR levels but also ensure a more uniform quality of service across the network.

Consequently, this research aims to fill the gap in the literature by proposing a novel solution that reconciles the need for low-latency communication with the demand for high signal quality, even in the most challenging edge-of-site scenarios.

III. PROPOSED METHOD

In this section, we propose an AP connection method for multi-homing that strikes a balance between suppressing the increase of FH cost and improving SINR.

A. CONCEPT OF MULTI-HOMING AP CONNECTION

Fig. 3 illustrates the concept of multi-homing AP connection and the main notations used in this section are listed in Table 2. APs located at the edge site boundaries can can perform multi-homing to a neighbouring edge site. Due to this multi-homing, vCPUs within the edge site can select APs from different edge sites to receive wireless signals. As shown in Fig. 3, an AP located at the boundary of Edge Site #1 is connected to the adjacent Edge Site #2, transmitting the duplicated wireless signals received from UEs. Through multi-homing, Edge Site #2, where the UE's vCPU is located, can receive wireless signals from the AP situated at Edge Site #1. Assuming AP 1 is additionally connected to Edge Site $Q_{\mathrm{MH},l}$ through multi-homing, the set of selectable APs for UE k expands to $\mathcal{L}'_{U_k} = \{l | \Omega_l = U_k \vee \Omega_{\mathrm{MH},l} = U_k\}$. This allows the selection of APs located in different edge sites as an AP cluster. This strategy facilitates the placement of vCPUs at edge sites and the formation of AP clusters that span across areas, which is expected to enhance the SINR. Furthermore, as depicted in Fig. 3, since UE#1 can utilize MEC, it not only enables the achievement of low latency but also allows for the cost-effective and straightforward duplication of radio signals. This duplication is facilitated by branching the signals output from wave division multiplexing (WDM) through optical couplers, which contributes to maintaining low latency without incurring significant additional costs. However, this method leads to an increase in the number of FHs due to the one-to-one connection between the AP and the edge site, resulting in a rise in FH costs. Furthermore, the need for multiple Q_l values necessitate the installation of FHs from AP l to multiple edge sites, which further escalates the FH costs. Therefore, there needs to be a balance between

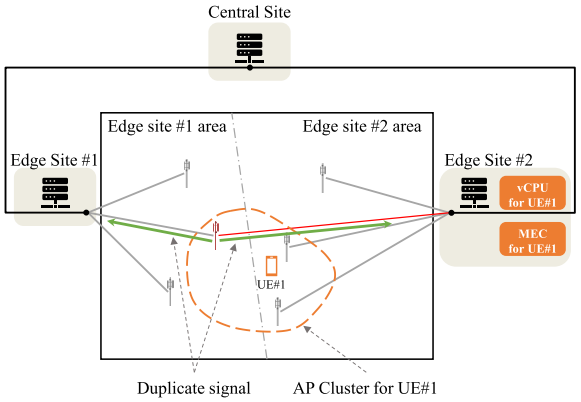


FIGURE 3. Concept of multi-home AP connection in FH for large-scale CF-mMIMO.

suppressing the increase in FH cost and improving SINR. To address this issue, it is necessary to reduce the number of FHs while maintaining a high SINR.

B. PROPOSED MULTI-HOME CONNECTION

In the proposed method, a heuristic algorithm is introduced that identifies AP interconnections likely to benefit from multi-homing due to their proximity to edge sites, resulting in fewer FH deployments. By prioritizing the connection of APs that are closer to the edge sites for multi-homing, only the necessary and sufficient radio signals for the users can be duplicated. This approach achieves a balance between cost and wireless quality. The flow of the proposed method is depicted in Fig. 4. In this method, the set of APs for multi-homing, $MH_{\mathrm{AP},l}$, and the pairs for connecting the FH, P_L, are stored at each step. Additionally, a threshold distance x_{th} meters is set between APs, and if this distance is not met, the process either reverts to STEP 1 or the selection of multi-homing APs is completed. Besides the distance threshold, there is an upper limit, $x_{limit-MH}$, on the number of APs that can be included in one multi-homing group. This procedure is repeated until the upper limit is reached, and if the limit is met, another multi-homing group is formed. Moreover, when selecting APs for multi-homing, an AP that has already been performing multi-homing is not used in another multi-homing group; this condition is applied to the multi-homing process.

Initially, in STEP 1 of Fig. 4, the pair of APs that are closest in terms of distance between them are identified. The selection of APs is made from a different edge site. At this time, $MH_{\mathrm{AP},l}$ will be [AP#1, AP#4], and P_l will be [(AP#1, AP#4)].

Subsequently, in STEP 2 of Fig. 4, the closest AP from within the APs in $MH_{\mathrm{AP},l}$ is selected, in a similar manner to STEP 1, but from a different area. At this time, APs that have already been multi-homing, such as AP#1 and AP#4 in this case, are excluded from the selection as they have already been determined as multi-homing. Distances between AP#1-AP#5, AP#1-AP#6, AP#1-AP#9, AP#4-AP#2, AP#4-AP#6,

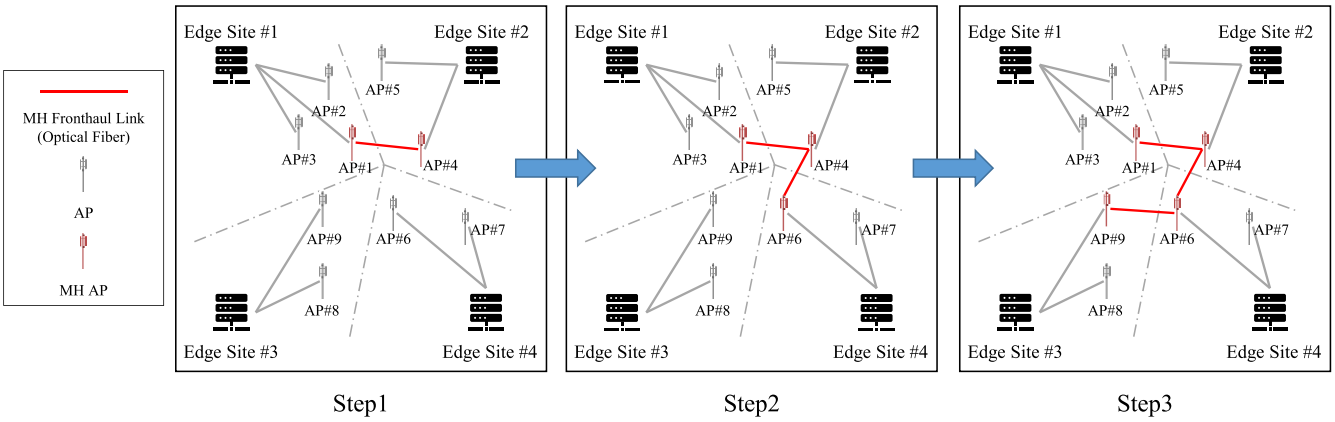


FIGURE 4. Proposed AP connection link design flow.

and AP#4-AP#9 are compared to find the combination with the shortest distance. At this time, if the distance between AP#4 and AP#6 is the shortest, then $MH_{\mathrm{AP},l}$ becomes [AP#1, AP#4, AP#6], and P_l is established as [(AP#1, AP#4), (AP#4, AP#6)].

Then, in STEP 3 of Fig. 4, APs are selected in the same manner as STEP 2. In this instance, since [AP#1, AP#4, AP#6] have already been determined as multi-homing, distances between pairs that do not connect these APs are compared. In this example, AP#6-AP#9 is the closest pair, then MH_{AP_l} becomes [AP#1, AP#4, AP#6, AP#9], and P_l is established as [(AP#1, AP#4), (AP#4, AP#6), (AP#6, AP#9)]. This process is repeated until the maximum number of APs able to perform multi-homing in one iteration, $x_{limit-MH}$, is reached. If $x_{limit-MH} = 4$, the process concludes at STEP 3 and returns to STEP 1, where multi-homing is created using APs other than the already determined multi-homing APs [AP#1, AP#4, AP#6, AP#9]. Additionally, these multi-homing APs can be assigned to all APs encompassing edge sites #1 to #4, making these APs available for AP cluster selection across edge sites #1 to #4. This facilitates the formation of AP clusters that span site boundaries. In this instance, it is assumed that the distance threshold x_{th} is not met during the steps. However, if the threshold is reached at any step, the multi-homing process concludes at that point and returns to STEP 1. Furthermore, if the threshold of x_{th} is met at STEP 1, the process of multi-homing formation itself is terminated.

Next, the proposed AP selection process is shown in detail in Algorithm 1.

x_{AP} represents the coordinates of the AP, x_{th} is the threshold distance between APs for multi-homing, $x_{limit-MH}$ is the maximum number of APs per multi-homing group, and Q_L denotes the edge site corresponding to AP l.

In steps 1-3, the L1 distance is used to calculate the distances from AP l to all other APs. Step 5 uses centroid conditions to terminate the multi-homing formation. The termination is triggered if all APs within AP_{dist} belong to the same edge site, or if the array size of AP_{dist} (the number of remaining APs) is either 0 or 1. This indicates that no additional multi-homing can be created, thus ending the multi-homing formation process. Subsequently, in steps 7-9, the closest AP to AP l, referred to as *targetAP*, is retained from the distances to all APs ($AP_{dist,l}$), specifically those APs that are situated in a different edge site from AP l. In step 10, *targetAP* is sorted in ascending order. By rearranging them in this way, it becomes easier to extract the ones with shorter distances. In step 12, one of the APs, which is proximate in distance to AP l, is incorporated into $MH_{\mathrm{AP},l}$ and initializes $MH_{\mathrm{AP},l}$. In steps 13-21, the creation of multi-homing pairs is undertaken until either the $x_{limit-MH}$ limit is met or no further multi-homing can be performed. This process mirrors the one outlined in steps 1 to 3 in Fig. 4. In step 14, a nearby AP from a different edge site from within $MH_{\mathrm{AP},l}$ is selected. In steps 15-16, if the distance to the selected AP satisfies x_{th}, we proceed to the next step; if it surpasses x_{th}, the multi-homing formation is terminated. In steps 17-18, APs meeting the stipulated conditions are added to $MH_{\mathrm{AP},l}$, and the AP with the smallest distance assessment is preserved as a pair with $MH_{\mathrm{AP},l}$ in P_l. It should be noted that $MH_{\mathrm{AP},l}$ retains the indices of the APs that will be multi-homing simultaneously, while P_l stores all the connection information for the APs that make up $MH_{\mathrm{AP},l}$. $MH_{\mathrm{AP},l}$ does not contain duplicate APs, however, in P_l, there exists a potential for AP duplication within the array. In step 22, all data of APs in $MH_{\mathrm{AP},l}$ are removed from AP_{dist}, facilitating the management of the remaining APs in the array. In steps 23-26, the sites for multi-homing are determined. In step 23, a unique function which eliminates duplicate elements from an array is utilized to create an array of edge sites to which each AP in $MH_{\mathrm{AP},l}$ is affiliated. This also represents the group of edge sites associated with $MH_{\mathrm{AP},l}$. In steps 24-26, for each AP in $MH_{\mathrm{AP},l}$, a unique function is employed to maintain the original edge site while incorporating additional sites for multi-homing. This method enables the assignment of multiple edge sites to the multi-homing APs.

Subsequently, the process flow described thus far is depicted in the flowchart in Fig. 5.

Algorithm 1 Proposed AP Selection

Input: x_{AP}, x_{th}, $x_{limit-MH}$, Q_L
1: **for** $l = 1$ to L **do** #Calc Distances to all APs from AP l
2: $AP_{dist,l} = \left\| \boldsymbol{x}_{AP,l} - \boldsymbol{x}_{AP,L} \right\|_1$
3: **end for**
4: **while true** # make multi-homing
5: **if** AP_{dist} is all same site | $length\left(AP_{dist,l}\right) == 0 or 1$
6: **break** # end of make multi-homing
7: **for** $l = 1$ to $length\left(AP_{dist,l}\right)$ **do**
8: $targetAP_{dist,l} = min\left[AP_{dist,l}\left(Q_l \neq Q_L\right)\right]$
9: **end for**
10: $targetAP_{dist} = sort\left(targetAP_{dist}\right)$ #sort ascend distance
11: **for** $l = targetAP_{dist}$ **do**
12: $MH_{\mathrm{AP},l} = [\mathrm{AP}\ l]$
13: **while true**
14: $APs =$ AP with the minimum dist within $MH_{\mathrm{AP},l}$
15: **if** $targetAP_{dist}\ (l, APs) \geq x_{th}$
16: ***break***
17: $MH_{\mathrm{AP}l} = [MH_{\mathrm{AP},l}, APs]$
18: $P_l = [P_l, [AP, MH_{\mathrm{AP},l}$ in the minimum for $APs]]$
19: **if** $length\left(MH_{AP_l}\right) == x_{limit-MH}$ | do not make multi-homing
20: **break**
21: **end while**
22: $AP_{dist}\left(MH_{\mathrm{AP},l}\right) = []$ # remove $MH_{\mathrm{AP},l}$
23: $MH_{site} =$ unique $\left[Q\left(MH_{\mathrm{AP},l}\right)\right]$
24: **for** $MH_l = MH_{\mathrm{AP},l}$ to **do**
25: $Q\left(MH_l\right) = \left[Q\left(MH_l\right),\ \text{unique}\left(MH_{site}, Q\left(MH_l\right)\right)\right]$
26: **end for**
27: **end for**
28: **end while**
29: **function** unique
30: Remove duplicate elements from an array
31: **end function**
Output: MH_{AP}, P_L

C. COST ANALYSIS FOR MULTI-HOME CONNECTION

In this study, we conduct a comparative analysis of the FH cost associated with each connection method, namely without MH (w/o MH), all AP MH (all MH), AP cluster base MH, random MH, and the proposed MH (PMH). Specifically, the FH cost of the random MH is calculated based on PMH, and the AP cluster base MH is calculated based on all MH. Taking into account the installation cost of APs CO_{AP}, the cost of the FH link between the AP and edge site CO_{FH}, and the cost of the (WDM) connection component, treated as a connection component between the AP and edge site CO_{WDM} [39], [40], are part of FH cost between the AP and edge site. Focusing on the installation cost of APs, $CO_{\mathrm{AP}} = C_{\mathrm{AP}}L$, and since the number of APs remains constant across each method, CO_{AP} is consistently the same across each method. Here, C_{AP} represents the installation cost coefficient of a single AP.

CO_{FH} is defined as:

$$CO_{FH} = \begin{cases} C_{\mathrm{fiber}} \Sigma_{l \in \mathcal{L}} \left\| \boldsymbol{x}_{\mathrm{AP},l} - \boldsymbol{x}_{\mathrm{site},Q_l} \right\|_1, & \text{w/o MH} \\ C_{\mathrm{fiber}} \{ \Sigma_{l \in \mathcal{L}} \left\| \boldsymbol{x}_{\mathrm{AP},l} - \boldsymbol{x}_{\mathrm{site},Q_l} \right\|_1 \\ \quad + \Sigma_{l \in \mathcal{L}_{\mathrm{MH}}} \left\| \boldsymbol{x}_{\mathrm{AP},l} - \boldsymbol{x}_{\mathrm{site},Q_{\mathrm{MH},l}} \right\|_1 \}, & all\ \text{MH} \\ C_{\mathrm{fiber}} \{ \Sigma_{l \in \mathcal{L}_{MH}} \left\| \boldsymbol{x}_{\mathrm{AP},l} - \boldsymbol{x}_{\mathrm{site},Q_l} \right\|_1 \\ \quad + \Sigma_{l \in \mathcal{L}_{MH}} \left\| P_{l,1} - P_{l,2} \right\|_1 \}, & \text{PMH}, \end{cases} \quad (6)$$

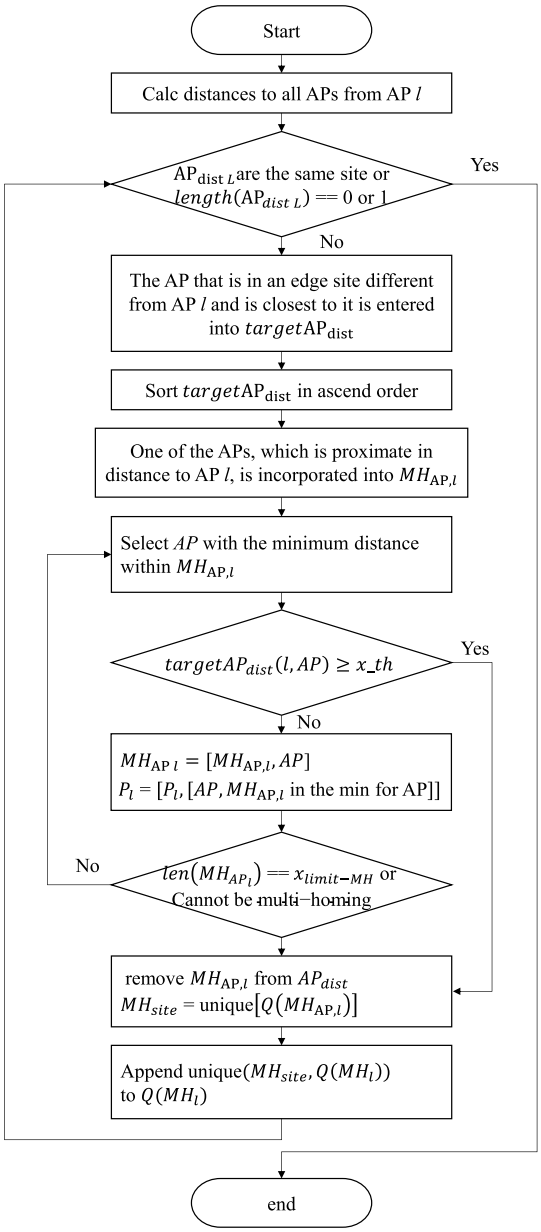


FIGURE 5. Flowchart for the Proposed AP Selection.

where, $\mathrm{C}_{\mathrm{fiber}}$ is the cost coefficient per 1 km of optical fiber used in FH, and $\mathcal{L}_{\mathrm{MH}}$ denotes the set of APs that have been additionally connected to the edge site via multi-homing. As illustrated in equation (6), the FH link cost across all methods increases as the distance between the connected APs increases.

CO_{WDM} is defined as:

$$CO_{\mathrm{WDM}} = \begin{cases} 2\Sigma_{l=1}^{L} C_{\mathrm{WDM},l}, & \text{w/o MH} \\ 2(\Sigma_{l=1}^{L} C_{\mathrm{WDM},l} + \Sigma_{l \in \mathcal{L}_{\mathrm{MH}}} C_{\mathrm{WDM},l}), & \text{all MH} \\ \Sigma_{l=1}^{L} C_{\mathrm{WDM},l} \left[2 \left| MH_{\mathrm{AP},l} \right| + 2 \left(\left| MH_{\mathrm{AP},l} \right| - 1 \right) \right], & \text{PMH} \end{cases} \quad (7)$$

where, $C_{\mathrm{WDM},l}$ is the cost coefficient for the WDM component corresponding to the transmission load $R_{\mathrm{FH},l}$, which is the radio signal transmitted by AP l. It is assumed that WDM components include WDM transmitters and receivers, muxponders and demuxponders, as well as transponders, and that the cost coefficient varies depending on the transmission load required for each FH. Each cost coefficient uses the values in [39], [40]. The transmission load in FH links is calculated based on the bandwidth and the quantization bit number of the radio signal [41].

IV. EVALUATION

In this section, we present the simulation conditions and the numerical results of the computer simulation to demonstrate the performance of the proposed method.

A. EVALUATION CONDITIONS

Table 3 outlines the conditions for the computer simulation, and Fig. 6 depicts the evaluation environment. UEs and APs are randomly distributed within a 1 km^2 area. Based on the user-centric RAN architecture, a single central site covers the entire area, and $J = 9$ edge sites are established to cover each subdivided area. The division of edge sites is accomplished using Voronoi regions [44], as depicted in Fig. 6, with each edge site randomly positioned within its respective area. Each AP is connected to the edge site of its corresponding divided area. Given that the terminal remains stationary, the calculation is executed ten times to enable a robust statistical evaluation. For this study, we adopt the following AP cluster formation algorithm proposed in [6], which has been referenced in over 500 publications.

- When UE k initiates communication, it sends a request to the AP with the best channel status. The AP that receives the request becomes the primary AP for UE k. The primary AP services UE k and assigns the pilot with the least pilot contamination to UE k at that time.
- For all APs that can be selected for the AP cluster for UE k, APs are associated with the cluster if the difference in channel gain between the primary AP of UE k and the AP falls within x dB.

In this paper, we set $x = 15$ for the current simulation.

B. EVALUATION METHOD

To confirm the effectiveness of the proposed methods, we compare the following five methods.

1. Proposed MH: This method involves the implementation of the advanced AP connectivity method as delineated in the proposal.
2. W/o MH: This scenario does not employ multi-homing, thereby limiting AP clusters to the range of APs connected to a single edge site.
3. All AP MH: In this configuration, every AP is multi-homing to all edge sites, thereby simulating an ideal cell-free environment.
4. AP cluster base MH: Initially, AP clusters are established for a given UE drop pattern, with APs spanning multiple sites multi-homing to the corresponding UE's edge site. This multi-homing configuration is then reused throughout the simulation.
5. Random MH: This method adheres to an equivalent quantity of multi-homing connections as the proposed method, with the interconnections between APs being selected at random.

TABLE 3. Simulation parameters.

Parameter	Value
The number of APs	100 to 500
The number of UEs	50 to 150
Location of APs and UEs	Uniform distribution
User traffic	Full buffer
Carrier frequency, Bandwidth	3.5 GHz, 100 MHz
Transmission power of UE	100 mW
Subcarrier spacing	30 kHz
Number of pilot sequences	24
Quantization bit rate	16 bits/symbol
Path loss and Channel fading	3GPP-UMi [42], Rayleigh fading
Channel estimation	MMSE [43]
Shadowing deviation	10 dB
Noise figure	7 dB

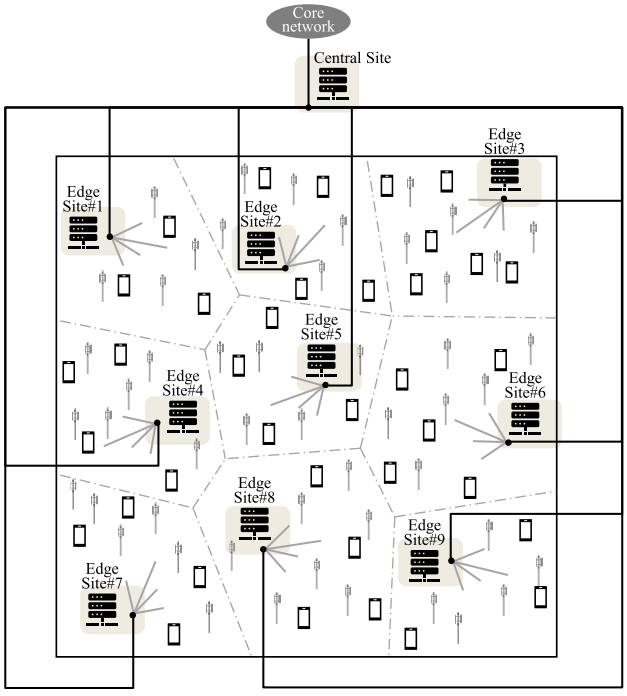


FIGURE 6. Evaluation environment. The figure shows that the number of edge sites is nine.

C. THROUGHPUT SIMULATION AND RESULTS

Fig. 7 depicts the 5%-tile uplink user throughput for the number of dropped UEs, while maintaining the number of APs at a fixed count of 400. With all AP MH, each AP is

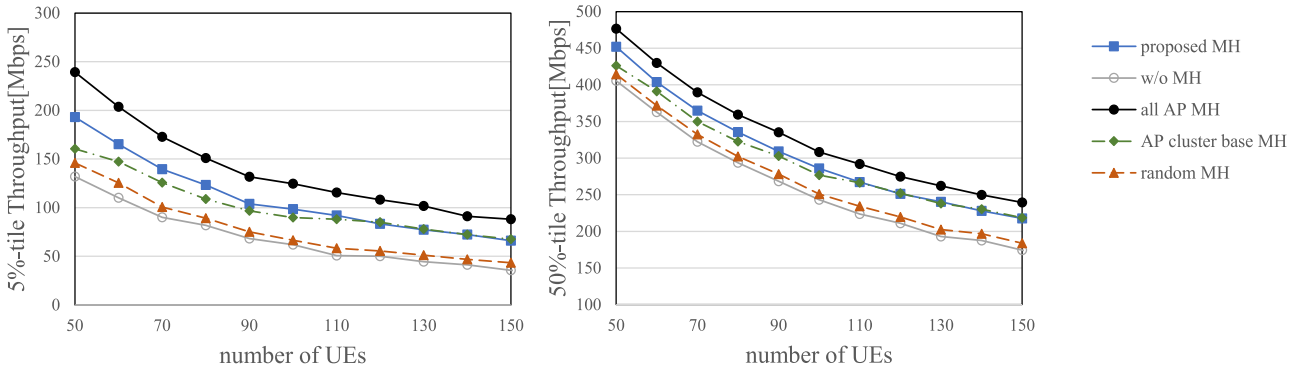


FIGURE 7. Comparison of the 5%,50%-tile user throughput for the number of UEs in the service area.

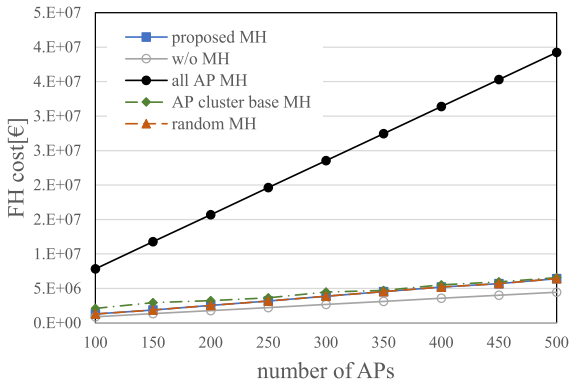


FIGURE 8. Comparison of the FH cost for the number of APs in the service area.

multi-homing to every site, enabling any UE to form an AP cluster that spans multiple sites, thereby leading to the highest throughput. This is due to all APs having the ability to form what is considered an ideal AP cluster.

The proposed MH, although not as high-performing as all AP MH, still achieves a higher throughput than other methods. This is because multi-homing APs near site boundaries facilitate UEs in proximity to these boundaries to form AP clusters that span sites, thus approximating an ideal cluster configuration. Additionally, the proposed MH focuses on APs that are close to site boundaries and are more likely to enhance the quality at these boundaries, resulting in values that are closest to all AP MH.

AP cluster base MH is the most efficient form of multi-homing in a snapshot environment. However, because the multi-homing APs are fixed, in this simulation with a variety of UE drops, it does not always select the most efficient multi-homing APs, leading to inferior performance compared to the proposed method.

As for random MH, it does not facilitate selection of the most efficient multi-homing APs that can contribute to forming AP clusters across site boundaries. As a result, while it does achieve higher throughput than scenarios w/o MH, it yields lower throughput compared to the other methods.

In case of the scenario w/o MH, the throughput is the lowest among all methods because it cannot form extensive AP clusters that span multiple sites. This particularly affects users near the area boundaries, which becomes more pronounced in the 5%-tile throughput.

D. FH COST ESTIMATION RESULTS

We evaluate the FH cost associated with each connection method. Fig. 8 depicts the cost for the number of dropped APs, while keeping the number of UEs at a fixed count of 100. Initially, as the number of APs remains constant across all patterns, there is no difference in the installation cost of APs. Moreover, as indicated in equation (6), the FH link cost for multi-homing escalates as the distance between APs increases.

Upon examining the results for all AP MH, the cost appears significantly higher compared to other methods. As demonstrated in equation (7), this is because all APs are drawing FH to all edge sites, resulting in a substantial number of WDM components.

With all AP MH, each AP is multi-homing to every site, leading to an increase in the number of FH connections between APs and edge sites. Consequently, significant amounts of fiber and WDM capacity become necessary, resulting in the highest cost among all methods.

The proposed MH has a lower cost, second only to the scenario w/o MH. This is because it performs multi-homing with APs located near site boundaries, which reduces the number of APs required for multi-homing. Unlike all AP MH, each multi-homing AP is connected to only one edge site, allowing for cost savings despite connections to multiple edge sites. Compared to w/o MH, the increase in cost is only due to the additional inter-AP connections.

AP cluster base MH incurs higher costs, second to all AP MH. Similar to all AP MH, it establishes FH connections from each AP to the sites involved in multi-homing. Thus, as the number of multi-homing APs increases, the cost escalates. It is less expensive than all AP MH because fewer APs are multi-homing, and it is more costly than the proposed method because each multi-homing AP adds FH connections to the sites it multi-homing to.

Random MH costs are nearly equivalent to the proposed method because the number of APs undergoing multi-homing is the same. Although the actual APs selected for multi-homing differ, this result suggests that the cost due to fiber distance is not as dominant as the cost of WDM, leading to a minimal difference.

W/o MH, the cost is the lowest since there are only FH connections between APs and edge sites, with no additional multi-homing infrastructure.

E. THROUGHPUT AND COST ASSOCIATED WITH MH APS

Fig. 9 and Fig. 10 depict the results when the number of APs is fixed at 400 and the number of UEs at 100, with variations in the number of APs undergoing multi-homing. The proposed MH varies the number of multi-homing APs by adjusting the value of x_{th}. For random MH, the number of multi-homing APs is the same as in the proposed MH, resulting in an identical count of APs. AP cluster base MH varies the number of multi-homing APs by adjusting the threshold for AP cluster formation during multi-homing. As for all AP MH, since all APs are always multi-homing, there is no variation in the number of multi-homing APs, and thus no fluctuation in results due to changes in multi-homing AP count.

All AP MH methods results in the highest throughput and cost. The high throughput is attributed to each AP being multi-homing to every site, enabling all UE to form ideal AP clusters. The cost is the highest because all APs are multi-homing to every site, which increases the number of FH connections required, and consequently the overall cost.

The proposed MH delivers a throughput that is lower than all AP MH but higher than other methods. Moreover, as the number of multi-homing APs increases, so does the throughput. This is because an increase in multi-homing APs leads to more multi-homing near site boundaries, making it easier to form AP clusters in these areas and thus approximating the ideal AP cluster configuration. In terms of cost, it trends at the lowest values. This is because multi-homing APs are connected to only one edge site and there is a cap on the number of APs selected for multi-homing, which helps contain costs.

AP cluster base MH is the most efficient in terms of AP clusters in a snapshot environment. However, because the multi-homing APs are fixed, it does not always select the most efficient multi-homing APs, resulting in performance that is inferior to the proposed MH. Similar to the proposed multi-homing, increasing the number of multi-homing APs leads to more edge sites per AP being multi-homing, which facilitates the formation of AP clusters across sites and increases throughput. In terms of cost, similar to throughput, as the number of multi-homing APs increases, so does the cost, due to the increase in FH connections per edge site.

Random MH performs worse than any other method. This is because the increase in the number of multi-homing APs does not effectively improve quality, as it does not facilitate selection of the most efficient multi-homing APs that can contribute to forming AP clusters across site boundaries. Regarding cost, it is nearly equivalent to or higher than the proposed MH. The increase in cost is due to the random selection of multi-homing APs, which may result in longer FH fiber distances compared to strategic multi-homing near site boundaries in the proposed MH.

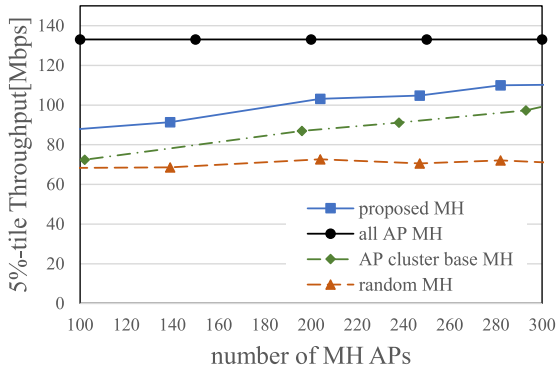


FIGURE 9. Comparison of the 5%-tile user throughput for the number of MH APs in the service area.

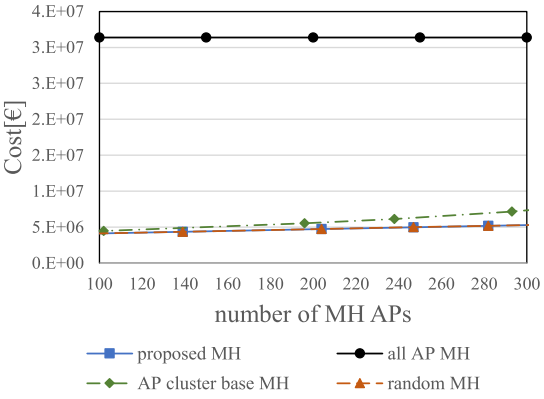


FIGURE 10. Comparison of the Cost for the number of MH APs in the service area.

As these results indicate, increasing the number of multi-homing APs enhances throughput but also leads to higher costs. Therefore, operators need to determine the number of multi-homing APs while considering the cost-effectiveness relative to the service quality they wish to provide in the area.

Future research directions include refining these approaches with advanced machine learning algorithms to dynamically optimize AP connections. Furthermore, emerging 6G technologies like reconfigurable intelligent surfaces and mobile fronthaul solutions are utilized for enhanced network performance, and prioritizing energy efficiency to develop sustainable networks. Considering standardization, the multi-homing technique, which is proposed in this paper, aims to improve radio quality and network efficiency in 6G networks. This is also connected to the O-RAN standard specification. The fronthaul specification, which belong to O-RAN standard, is called Split Option 7-2x. This specification defines a functional split between distributed units and radio units. The transmission of the data between these nodes is accomplished

via an ethernet frame, which is compatible with our proposed multi-homing scheme.

V. CONCLUSION

In this paper, we have presented a comprehensive study on the deployment of vCPUs within a CF-mMIMO architecture, aimed at addressing the stringent requirements of the forthcoming 6G networks. Our focus has been on overcoming the challenges associated with ensuring low latency and high radio quality across distributed and hierarchical network structures, which are critical for the support of latency-sensitive and bandwidth-intensive applications.

The degradation of radio quality at the edges of distributed sites is identified as a significant challenge, primarily due to inter-site interference and the limitations in AP selection. To mitigate these issues, a multi-homing technique is proposed that allows APs to connect to multiple edge sites. This approach not only improves radio quality for users located at the site boundaries but also maintains the network's cost-effectiveness by minimizing the need for additional FH installations. Through our heuristic algorithm, we demonstrated that strategic multi-homing can effectively balance the trade-off between FH costs and radio quality. The simulation results confirm that the proposed method approaches the ideal throughput values while keeping the FH costs in check.

REFERENCES

[1] H. Tataria, M. Shafi, A. F. Molisch, M. Dohler, H. Sjöland, and F. Tufvesson, "6G wireless systems: Vision, requirements, challenges, insights, and opportunities," *Proc. IEEE*, vol. 109, no. 7, pp. 1166–1199, Jul. 2021.

[2] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO: Uniformly great service for everyone," in *Proc. IEEE 16th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Stockholm, Sweden, Jun. 2015, pp. 201–205.

[3] S. Buzzi and C. D'Andrea, "Cell-free massive MIMO: User-centric approach," *IEEE Wireless Commun. Lett.*, vol. 6, no. 6, pp. 706–709, Dec. 2017.

[4] C. D'Andrea and E. G. Larsson, "User association in scalable cell-free massive MIMO systems," in *Proc. 54th Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, USA, Nov. 2020, pp. 826–830.

[5] T. Murakami, N. Aihara, A. Ikami, Y. Tsukamoto, and H. Shinbo, "Analysis of CPU placement of cell-free massive MIMO for user-centric RAN," in *Proc. NOMS-IEEE/IFIP Netw. Oper. Manage. Symp.*, Budapest, Hungary, Apr. 2022, pp. 1–7.

[6] E. Björnson and L. Sanguinetti, "Scalable cell-free massive MIMO systems," *IEEE Trans. Commun.*, vol. 68, no. 7, pp. 4247–4261, Jul. 2020.

[7] Z. Liu, J. Zhang, Y. Li, and Y. Ji, "Hierarchical MEC servers deployment and user-MEC server association in C-RANs over WDM ring networks," *Sensors*, vol. 20, no. 5, p. 1282, Feb. 2020.

[8] P. E. Engelstad, A. Tonnesen, A. Hafslund, and G. Egeland, "Internet connectivity for multi-homed proactive ad hoc networks," in *Proc. IEEE Int. Conf. Commun.*, Paris, France, Jun. 2004, pp. 4050–4056.

[9] S. Kashihara, T. Nishiyama, K. Iida, H. Koga, Y. Kadobayashi, and S. Yamaguchi, "Path selection using active measurement in multi-homed wireless networks," in *Proc. Can. Conf. Electr. Comput. Eng., Toward Caring Humane Technol.*, 2004, pp. 273–276.

[10] M. Peng, Y. Sun, X. Li, Z. Mao, and C. Wang, "Recent advances in cloud radio access networks: System architectures, key techniques, and open issues," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 2282–2308, 3rd Quart., 2016.

[11] P. Chanclou, H. Suzuki, J. Wang, Y. Ma, M. R. Boldi, K. Tanaka, S. Hong, C. Rodrigues, L. A. Neto, and J. Ming, "How does passive optical network tackle radio access network evolution?" *J. Opt. Commun. Netw.*, vol. 9, no. 11, pp. 1030–1040, Nov. 2017.

[12] A. Fernandez and N. Stol, "OPEX simulation study of PONs based on network geometric and Markov cost models," in *Proc. Int. Conf. Opt. Netw. Design Modeling*, Stockholm, Sweden, May 2014, pp. 252–257.

[13] J. Zhang, E. Björnson, M. Matthaiou, D. W. K. Ng, H. Yang, and D. J. Love, "Prospective multiple antenna technologies for beyond 5G," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1637–1660, Aug. 2020.

[14] S. Buzzi, C. D'Andrea, A. Zappone, and C. D'Elia, "User-centric 5G cellular networks: Resource allocation and comparison with the cell-free massive MIMO approach," *IEEE Trans. Wireless Commun.*, vol. 19, no. 2, pp. 1250–1264, Feb. 2020.

[15] A. Checko, H. L. Christiansen, Y. Yan, L. Scolari, G. Kardaras, M. S. Berger, and L. Dittmann, "Cloud RAN for mobile networks—A technology overview," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 1, pp. 405–426, 1st Quart., 2015.

[16] R. W. Heath Jr., and A. Lozano, *Foundations of MIMO Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2018.

[17] H. A. Ammar and R. Adve, "Power delay profile in coordinated distributed networks: User-centric V/S disjoint clustering," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Nov. 2019, pp. 1–5.

[18] D. Liu, S. Han, C. Yang, and Q. Zhang, "Semi-dynamic user-specific clustering for downlink cloud radio access network," *IEEE Trans. Veh. Technol.*, vol. 65, no. 4, pp. 2063–2077, Apr. 2016.

[19] Y. Tsukamoto, A. Ikami, N. Aihara, T. Murakami, H. Shinbo, and Y. Amano, "User-centric AP clustering with deep reinforcement learning for cell-free massive MIMO," in *Proc. Int. ACM Symp. Mobility Manage. Wireless Access*, Oct. 2023, pp. 17–24.

[20] Y. Al-Eryani, M. Akrout, and E. Hossain, "Multiple access in cell-free networks: Outage performance, dynamic clustering, and deep reinforcement learning-based design," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 4, pp. 1028–1042, Apr. 2021.

[21] H. A. Ammar, R. Adve, S. Shahbazpanahi, G. Boudreau, and K. V. Srinivas, "Downlink resource allocation in multiuser cell-free MIMO networks with user-centric clustering," *IEEE Trans. Wireless Commun.*, early access, Mar. 19, 2022, doi: 10.1109/TWC.2021.3104456.

[22] H. A. Ammar, R. Adve, S. Shahbazpanahi, G. Boudreau, and K. V. Srinivas, "Distributed resource allocation optimization for user-centric cell-free MIMO networks," *IEEE Trans. Wireless Commun.*, early access, May 13, 2022, doi: 10.1109/TWC.2021.3118303.

[23] X. Chai, H. Gao, J. Sun, X. Su, T. Lv, and J. Zeng, "Reinforcement learning based antenna selection in user-centric massive MIMO," in *Proc. IEEE 91st Veh. Technol. Conf. (VTC-Spring)*, May 2020, pp. 1–6.

[24] L. Sun, J. Hou, and R. Chapman, "Multi-agent deep reinforcement learning for access point activation strategy in cell-free massive MIMO networks," in *Proc. IEEE Conf. Comput. Commun. Workshops*, May 2023, pp. 1–6.

[25] A. Ikami, Y. Tsukamoto, T. Murakami, H. Shinbo, and Y. Amano, "Distributed DRL with multiple learners for AP clustering in large-scale cell-free deployment," in *Proc. IEEE 99th Veh. Technol. Conf.*, Singapore, Jun. 2024, pp. 1–6.

[26] A. Ikami, Y. Tsukamoto, N. Aihara, T. Murakami, and H. Shinbo, "Interference suppression for distributed CPU deployments in cell-free massive MIMO," in *Proc. IEEE 96th Veh. Technol. Conf. (VTC-Fall)*, London, U.K., Sep. 2022, pp. 1–6.

[27] A. Ikami, N. Aihara, Y. Tsukamoto, T. Murakami, and H. Shinbo, "Cooperation method between CPUs in large-scale cell-free massive MIMO for user-centric RAN," *IEEE Access*, vol. 11, pp. 95267–95277, 2023.

[28] I. Atzeni, B. Gouda, and A. Tölli, "Distributed precoding design via over-the-air signaling for cell-free massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 20, no. 2, pp. 1201–1216, Feb. 2021.

[29] J. Zhang, J. Zhang, E. Björnson, and B. Ai, "Local partial zero-forcing combining for cell-free massive MIMO systems," *IEEE Trans. Commun.*, vol. 69, no. 12, pp. 8459–8473, Dec. 2021.

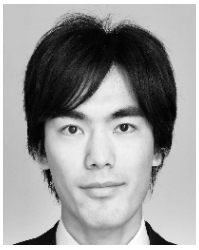
[30] E. Nayebi, A. Ashikhmin, T. L. Marzetta, H. Yang, and B. D. Rao, "Precoding and power optimization in cell-free massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, pp. 4445–4459, Jul. 2017.

[31] G. Interdonato, P. Frenger, and E. G. Larsson, "Scalability aspects of cell-free massive MIMO," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2019, pp. 1–6.

[32] J. Zhang, S. Chen, Y. Lin, J. Zheng, B. Ai, and L. Hanzo, "Cell-free massive MIMO: A new next-generation paradigm," *IEEE Access*, vol. 7, pp. 99878–99888, 2019.

[33] N. Wang, E. Hossain, and V. K. Bhargava, "Backhauling 5G small cells: A radio resource management perspective," *IEEE Wireless Commun.*, vol. 22, no. 5, pp. 41–49, Oct. 2015.

[34] U. Siddique, H. Tabassum, E. Hossain, and D. I. Kim, "Wireless backhauling of 5G small cells: Challenges and solution approaches," *IEEE Wireless Commun.*, vol. 22, no. 5, pp. 22–31, Oct. 2015.
[35] E. Sharifi, M. M. Feghhi, G. Azarnia, S. Nouri, D. Lee, and M. J. Piran, "Channel estimation based on compressed sensing for massive MIMO systems with lens antenna array," *IEEE Access*, vol. 11, pp. 79016–79032, 2023.
[36] M. S. Oh, A. B. Das, S. Hosseinalipour, T. Kim, D. J. Love, and C. G. Brinton, "A decentralized pilot assignment algorithm for scalable O-RAN cell-free massive MIMO," *IEEE J. Sel. Areas Commun.*, vol. 42, no. 2, pp. 373–388, Feb. 2024.
[37] Ö. T. Demir, M. Masoudi, E. Björnson, and C. Cavdar, "Cell-free massive MIMO in O-RAN: Energy-aware joint orchestration of cloud, fronthaul, and radio resources," *IEEE J. Sel. Areas Commun.*, vol. 42, no. 2, pp. 356–372, Feb. 2024.
[38] R. Beerten, V. Ranjbar, A. P. Guevara, and S. Pollin, "Cell-free massive MIMO in the O-RAN architecture: Cluster and handover strategies," in *Proc. IEEE Global Commun. Conf.*, Dec. 2023, pp. 5943–5948.
[39] F. Yaghoubi, M. Mahloo, L. Wosinska, P. Monti, F. de Souza Farias, J. C. W. A. Costa, and J. Chen, "A techno-economic framework for 5G transport networks," *IEEE Wireless Commun.*, vol. 25, no. 5, pp. 56–63, Oct. 2018.
[40] F. Rambach, B. Konrad, L. Dembeck, U. Gebhard, M. Gunkel, M. Quagliotti, L. Serra, and V. Lopez, "A multilayer cost model for metro/core networks," *J. Opt. Commun. Netw.*, vol. 5, no. 3, pp. 210–225, Mar. 2013.
[41] *Small Cell Virtualization: Functional Splits and Use Cases*, Small Cell Forum, U.K., Jan. 2016.
[42] *Study on Channel Model for Frequencies From 0.5 to 100 GHz*, document TR 38.901 V16.1.0, 3GPP, Dec. 2019.
[43] E. Björnson and L. Sanguinetti, "Making cell-free massive MIMO competitive with MMSE processing and centralized implementation," *IEEE Trans. Wireless Commun.*, vol. 19, no. 1, pp. 77–90, Jan. 2020.
[44] F. Aurenhammer, "Voronoi diagrams—A survey of a fundamental geometric data structure," *ACM Comput. Surveys*, vol. 23, no. 3, pp. 345–405, Sep. 1991.



AKIO IKAMI received the B.E. degree in electrical and electronic engineering from Kyoto University, Kyoto, Japan, in 2011, and the M.Sc. degree in communications and computer engineering from the Graduate School of Informatics, Kyoto University, in 2013. He joined KDDI Corporation, Tokyo, Japan, in 2013, and became engaged in the development of access networks. Since 2018, he has been with KDDI Research, Inc., Saitama, Japan. His research interests include RAN management and optimization for 6G. He received the Young Researcher's Award from IEICE in 2020 and the Best Paper Award at the IEEE VTC2022 Fall.

AMR AMRALLAH (Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical engineering from Aswan University, Aswan, Egypt, in 2007 and 2018, respectively, and the Ph.D. degree in electrical and electronic engineering from Tokyo Institute of Technology, Tokyo, Japan, in 2023. From 2008 to 2013, he was a Telecommunications Maintenance Engineer with Huawei Technologies Ltd., managed service projects in Southern Region of Egypt. From 2013 to 2019, he was a Senior Telecommunications Maintenance Engineer with Egyptian National Railways (ENR), Aswan. He has been a Researcher with KDDI Research, Inc., Saitama, Japan, since May 2023. His research interests include B5G/6G networks, UAV wireless communication networks, and RAN optimization and management.

YU TSUKAMOTO received the A.E. degree in electronic control system engineering from the National Institute of Technology, Numazu College, Shizuoka, Japan, in 2012, the B.E. degree in electrical engineering from Kyoto Institute of Technology, Kyoto, Japan, in 2014, and the M.E. degree in electrical engineering from Kyoto University, Kyoto, in 2016. He joined KDDI Corporation, Tokyo, Japan, in 2016. Since 2017, he has been a member with KDDI Research, Inc., Saitama, Japan, where he studies radio access network technologies.

TAKAHIDE MURAKAMI received the B.E., M.E., and Ph.D. degrees in communication engineering from Tohoku University, Sendai, Japan, in 2002, 2004, and 2007, respectively. He joined KDDI Corporation, in 2007, and is currently a member with KDDI Research, Inc., (formerly KDDI Research and Development Laboratories). His research interests include wireless communications and radio access networks.

HIROYUKI SHINBO received the B.S. degree in electro information communication and the M.S degree from the Graduate School of Information and System, The University of Electro-Communications, Tokyo, Japan, in 1987 and 1990, respectively. He joined KDD Corporation (now KDDI Corporation) in 1999 and was with KDD Research and Development Laboratories (now KDDI Research, Inc.). He was seconded to work with the Advanced Telecommunications Research Institute International, from 2013 to 2016. He is currently the Senior Manager of the Advanced Radio Application Laboratory, KDDI Research, Inc. His research interests include beyond 5G/6G systems (especially radio access networks), TCP/IP, flying base stations, network operation systems, and space communications.

YOSHIAKI AMANO received the B.E. and M.E. degrees in electrical and electronic engineering from Nagoya University, Aichi, Japan, in 1995 and 1999, respectively. He joined KDD Corporation (currently KDDI Corporation), in 1999, where he was involved in research and development on CDMA cellular systems, smart antenna, and wireless performance evaluation and improvement of mobile terminals. He also engaged with the Energy Business Planning Division, KDDI Corporation, to launch the electric power retail business, from 2015 to 2018. Since 2018, he launched research and development on dynamic spectrum sharing system and intelligent reflecting surface with KDDI Research, Inc. Before that, he has led research on wireless communication systems and radio access networks for beyond 5G/6G. He received the Young Researcher's Award from IEICE and the Meritorious Award on Radio from the Association of Radio Industries and Businesses (ARIB) in 2005 and 2011, respectively.