Geo-distributed and Dynamic Control Plane Management toward Green Mobile Networks

Masayuki Kurata, Akio Ikami, and Masaki Suzuki
KDDI Research, Inc., Saitama, Japan
{ma-kurata, ak-ikami, and mak-suzuki}@kddi.com

***Abstract*—The growing number of user equipment (UE) connections in mobile networks amplifies signaling messages (SMsgs) in the control plane (C-plane). Since a hierarchical site architecture is adopted, mobile networks strain site infrastructure and inter-site backhaul links, resulting in higher power consumption. To address this, we propose Dynamic Hybrid Distributed and Centralized C-plane Placement (D-HDCCP), which dynamically allocates UEs to geo-distributed C-plane sites based on mobility profiles. Specifically, moving UEs are allocated to central/regional C-plane sites to reduce total SMsgs during handovers, while stationary UEs are allocated to edge C-plane sites to minimize inter-site SMsgs for state management. When a mobility profile changes, D-HDCCP evaluates whether reallocation maximizes energy efficiency by an online, threshold-driven decision mechanism that considers resource constraints at edge sites. This assessment is critical when shifting from moving to stationary and facing resource limitations at edge sites, as improper migration can lead to energy inefficiencies. A handover-synchronized migration procedure minimizes SMsg overhead for reallocation if a migration is performed. Extensive simulations in urban commuter UE scenarios show that D-HDCCP, when fully deployed at edge sites, reduces power consumption by up to 11.0% during peak hours and 5.2% over a full day, compared to legacy C-plane placement.**

I. Introduction

With the rapid proliferation of the Internet of Things (IoT) and advancements in user equipment (UE) [1], the control plane (C-plane) of mobile networks, encompassing both the Radio Access Network (RAN) and Mobile Core Network (MCN), faces increasing threats from rising signaling load. This load demands significant network resources, such as processing power and bandwidth, which in turn raises power consumption [2]. Therefore, efficient management of signaling messages (SMsgs) exchanged between C-plane network function instances (NFIs) is crucial for green mobile networks.

This work targets total and inter-site SMsgs as key metrics contributing to the power consumption in the C-plane. These are observed in legacy mobile networks that adopt a three-level hierarchical site structure consisting of edge, regional, and central sites [3]. In this structure, RAN and MCN C-plane NFIs are deployed at regional and central sites, respectively. Reducing total SMsgs decreases throughput requirements in NFIs, contributing to energy savings in site infrastructures. Moreover, reducing inter-site SMsgs helps lower the non-negligible power consumption in inter-site backhaul links [4].

Existing efforts to reduce SMsgs include both standards-compliant and non-standards-compliant approaches. The former methods [5]–[8] are built on frameworks defined by standards organizations such as the Third Generation Partnership Project (3GPP) (e.g., [9]–[13]). Conversely, the latter methods [14]–[16] substantially reconfigure the established NF structure and communication protocols between NFIs. This deviation from standard practices creates substantial barriers to adoption in commercial environments and leads to a lack of backward compatibility with existing mobile networks [17].

Hybrid Distributed and Centralized C-plane Placement (HDCCP), as explored in [8], adopts a unique edge-centric approach explicitly designed within the 3GPP framework. HDCCP geographically deploys RAN and MCN C-plane NFIs based on the distribution of UE mobility profiles. Specifically, stationary UEs, such as fixed IoT sensors and residence UEs, are allocated to edge C-plane sites. This allocation reduces inter-site SMsgs for state management procedures by leveraging edge computing to eliminate forwarding SMsgs to central sites. Conversely, moving UEs, such as autonomous mobile robots and vehicles, are allocated to regional or central C-plane sites. This allocation minimizes total SMsgs by facilitating handovers that switch only a subset of NFIs, since the NF of RAN, which serves as an anchor, rarely changes.

However, HDCCP encounters challenges in dynamic mobility scenarios, where changes in mobility profiles can lead to inefficient SMsg management. Specifically, when UE mobility profiles shift from stationary to moving, continued allocation to edge C-plane sites leads to excessive SMsg overhead due to handover procedures that require switching multiple NFIs. Conversely, when UE mobility profiles shift from moving to stationary, continued allocation to central C-plane sites leads to excessive inter-site SMsgs for state management procedures. These limitations underscore the necessity of appropriate migration mechanisms for reallocation that respond to mobility profile transitions to save energy in the C-plane.

Building on these analyses, Dynamic HDCCP (**D-HDCCP**) is proposed as an extended framework incorporating an adaptive migration strategy. Energy-efficient migration is achieved through two key approaches: synchronizing migration with the handover procedure to minimize SMsgs for reallocation and employing an online, threshold-driven migration decision mechanism that accounts for site resource constraints to prevent unnecessary migrations. This cautious decision is made when UE mobility profiles shift from moving to stationary, with the threshold adjusted based on the UEs' procedural characteristics during the stationary period and the cumulative

ISBN 978-3-903176-72-0

distribution function (CDF) derived from their statistics.

Simulation-based analyses assess the effectiveness of D-HDCCP and its migration decision mechanism across different deployment phases for edge site computing resources. Moreover, simulations leveraging real-world C-plane and urban mobility data [18], [19] demonstrate that commuter UEs represent a well-suited use case. Both total and inter-site SMsgs are evaluated based on their impact on power consumption, providing a comprehensive assessment of energy efficiency—an essential requirement for mobile networks [1]. This evaluation refers to the methodology outlined in [4]. The results show that D-HDCCP reduces C-plane power consumption by 11.0% during peak hours and by 5.2% overall than legacy C-plane placement (LCP), assuming full deployment at edge sites.

This paper is organized as follows. Sec. II reviews existing efforts for SMsg reduction and HDCCP, which is static geo-distributed C-plane management. Sec. III presents the system model, formulates the objective, and details the migration procedure along with an online algorithm for energy-efficient migration. Sec. IV evaluates the effectiveness of D-HDCCP across different deployment phases in commuter UE scenarios and provides operational insights. Finally, Sec. V concludes.

II. Background

This work discusses assuming a 5G mobile network, the latest 3GPP standard [9]–[13]. First, we provide an overview of the 5G network architecture and C-plane procedures. Next, we review related works on SMsg reduction from 3GPP and non-3GPP perspectives. Finally, HDCCP [8] and its limitations in handling dynamic mobility profile changes are discussed.

A. 5G Mobile Network and Procedures

As shown in Fig. 1, RAN and MCN are decoupled into the C-plane and User plane (U-plane), with virtualized stateful NFIs deployed. In the C-plane, MCN NFIs use service-based architectures (SBA), while RAN and MCN NFIs communicate via point-to-point connections. Figure 2 depicts that C-plane RAN and MCN NFIs are typically deployed at regional and central sites. U-plane NFIs are assumed to be deployed at various sites with edge computing capabilities [1].

According to [18], mobile networks mainly observe the following state management and handover procedures.

Service Request and AN Release manage active and idle connections between the UE and RAN/MCN, ensuring fair utilization of UEs' wireless resources by frequently alternating between assigning and releasing them. The involved C-plane NFs are the Central Unit - Control Plane (CU-CP), Access and Mobility Management Function (AMF), and Session Management Function (SMF). Registration and Deregistration are required at the start and end of mobile communication. To handle these, authentication, policy control and data management NFs (other NFs in Fig. 1) are additionally used.

Handover procedure occurs when a UE moves to antenna sites under a different C-plane site. There are multiple handover options involving C-plane NFIs, executed in the following order if feasible: Intra CU Handover, where only RAN

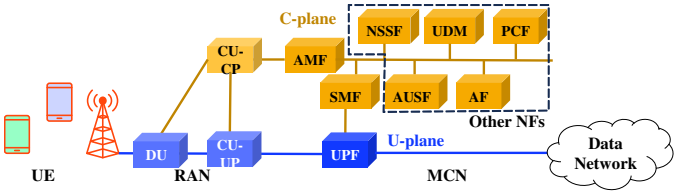


Fig. 1: 5G mobile networks which encompass RAN and MCN.

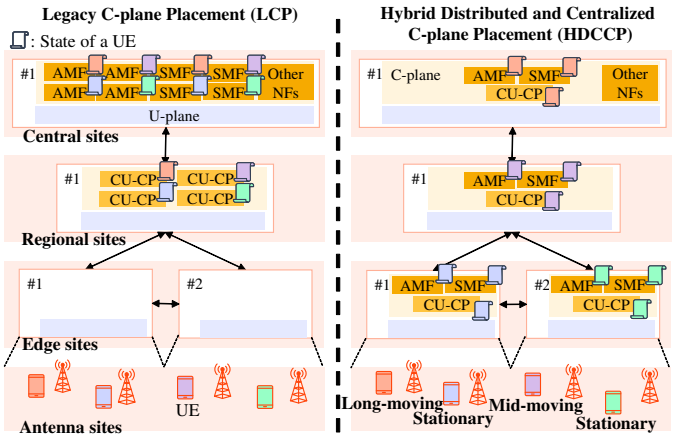


Fig. 2: Overview of LCP and HDCCP.

U-plane NFIs are switched, anchored by CU-CP; Intra AMF Handover, where RAN NFIs are switched, anchored by AMF; and Inter AMF Handover, involving the switching of RAN and MCN NFIs. These handover procedures transfer UE states between source and target C-plane NFIs. Switching more NFIs requires an increased number of SMsgs (see Table I).

B. Existing Work for Signaling Message Efficiency

Many studies on energy efficiency in mobile networks have primarily focused on the U-plane [20]. However, ensuring energy efficiency in the C-plane is equally critical, especially given the anticipated surge in simultaneous connections from heterogeneous IoT devices in future networks [1]. Moreover, IoT device communications typically generate similar traffic volumes in both the U-plane and C-plane, thereby exerting comparable impacts on each [21]. Thus, developing effective methods to mitigate the signaling load is imperative [2].

To realize SMsg reduction, approaches like L^2*5GC* [14], *Organic Core Network (OCN)* [15], and *Pp5gs* [16] introduce significant customizations to C-plane NFs. L^2*5GC* replaces HTTP with a shared memory-based protocol, reducing SMsg across network links. *OCN* and *Pp5gs* reconfigure NFs to minimize inter-NF SMsg exchanges, with *OCN* also integrating a web-friendly MCN architecture to enhance operational efficiency. However, these deviations from standard specifications add complexity and costs [17] in ensuring interoperability with existing deployments and other operators' mobile networks.

Therefore, maintaining compatibility with existing standards is also essential. The work in [5] introduces reactive methods for reducing SMsgs by dynamically reconfiguring specific NFI parameters, such as blocking selected procedures or adjusting request intervals for particular UEs. Similarly, the group handover mechanism described in [6] reduces SMsgs

TABLE I: Procedures that influence HDCCP arrangement.

	# of total SMsgs	# of inter-site SMsgs in LCP
Service Request	18	8
AN Release	12	5
Intra CU Handover	10	8
Intra AMF Handover	30	17
Inter AMF Handover	51	23

by executing handover procedures collectively for groups of authorized UEs, with the secure authentication mechanism extended within the 3GPP framework. Furthermore, the work in [7] enhances SMsg efficiency in UE discovery by integrating AI-based mobility prediction with conventional NFs.

Besides these works, HDCCP [8] offers a unique backward-compatible solution for SMsg reduction by integrating C-plane functionality at edge sites. The increasing adoption of edge computing [1] enhances its practicality and scalability.

C. Hybrid Distributed and Centralized C-plane Placement

Table I lists the procedures that influence the HDCCP arrangement. Due to the low occurrence rates of Registration (0.1%) and Deregistration (0.2%) [18] and the stringent data synchronization requirements, the distribution of other NFs in Fig. 1 is not performed. Instead, as shown in Fig. 2, HDCCP geographically distributes CU-CP, AMF, and SMF instances based on UE mobility profiles. The mobility profiles are classified to align with the levels of a hierarchical site structure (see Appendix for details). As a three-level hierarchy is targeted, they are categorized into three types: "stationary," "mid-moving," and "long-moving."

Figure 2 depicts UE allocation in HDCCP to reduce SMsgs. (i) **Stationary UE** is allocated to an edge C-plane site. This enables local processing of Service Request and AN Release without forwarding inter-site SMsgs to the regional and central sites. (ii) **Long-moving UE** is allocated to a central C-plane site. By extending the coverage of antenna sites managed by the CU-CP instance, the occurrence of Intra CU Handover is facilitated in handovers, thereby reducing both total and inter-site SMsgs. (iii) **Mid-moving UE**, which moves antenna sites controlled by a regional C-plane site, is allocated to a regional C-plane site, adopting a hybrid allocation rule between (i) and (ii). This facilitates Intra CU Handover during handovers and reduces inter-site SMsgs as much as possible by avoiding forwarding to central sites.

However, the static UE allocations in HDCCP can significantly increase SMsgs when mobility profiles transition. Specifically, when a UE mobility profile shifts from stationary to moving, continued UE allocation in the edge C-plane site results in Inter AMF Handover as the UE moves between antenna sites under different edge C-plane sites. This leads to a substantial increase in both total and inter-site SMsgs for each handover. On the other hand, when a UE mobility profile shifts from moving to stationary, continued UE allocation to the central C-plane site can cause excessive inter-site SMsgs due to procedures other than handovers. These issues highlight the necessity for HDCCP to dynamically reallocate UEs in response to transitioning mobility profiles. Furthermore, resource constraints at edge sites exacerbate this challenge by hindering the migration of UEs that would benefit from it.

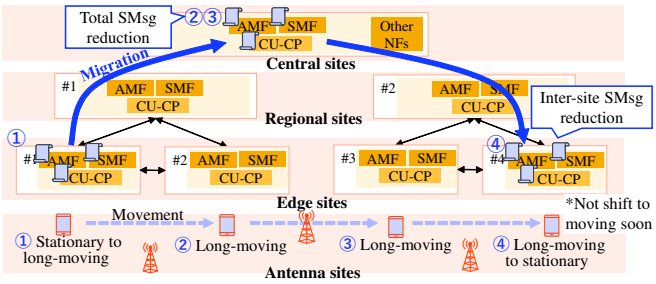


Fig. 3: Example of UE management by D-HDCCP with an adaptive migration strategy as the mobility profile transitions.

III. Dynamic HDCCP (D-HDCCP)

To address the SMsg inefficiency caused by dynamic mobility profile changes, Dynamic HDCCP (D-HDCCP) with an adaptive migration strategy for UE reallocation is proposed. As shown in Fig. 3, when a UE mobility profile shifts from stationary to moving, UE is migrated to a central or regional C-plane site, thereby preventing handover procedures with numerous SMsgs, i.e., Inter AMF Handover. Conversely, when the UE mobility profile shifts from moving to stationary, D-HDCCP carefully assesses its benefit before initiating it. This cautious approach accounts for scenarios where the mobility profile may quickly revert to moving after migration, and another migration is required. In such cases, the SMsgs generated by successive migrations could outweigh its benefits.

This section presents its overview and establishes its formulated problem as NP-hard. To solve this, an online threshold-driven algorithm is developed to greedily determine whether migration should be triggered when a UE's mobility profile transitions, using an SMsg-efficient migration procedure synchronized with a handover procedure. Furthermore, commuter UEs are identified as a target use case.

A. System Assumptions and Management Plane Design

As illustrated in Fig. 4, commodity servers with symmetric hardware are assumed to be deployed at site infrastructures to host C-plane NFIs. In the management plane [13], log data generated by servers and NFIs is aggregated and analyzed to extract UE and network characteristics. These analyses encompass real-time monitoring of requested procedures [22], UE mobility prediction, and attribute-based categorization (e.g., commuters and residents) [23]. To support an energy-efficient migration by D-HDCCP, the cumulative distribution function (CDF) $F(x)$ is statistically computed using log data to represent the frequency of state management procedures during stationary periods at each antenna site. Additionally, resource usage at each C-plane site is continuously monitored.

Mobility profiles are derived from UE handover occurrence logs and the associated antenna sites. A UE is classified as stationary if no handover logs are detected for a certain period. If handover logs are only observed from antenna sites managed by a specific regional C-plane site, the UE is categorized as mid-moving; otherwise, it is classified as long-moving. The

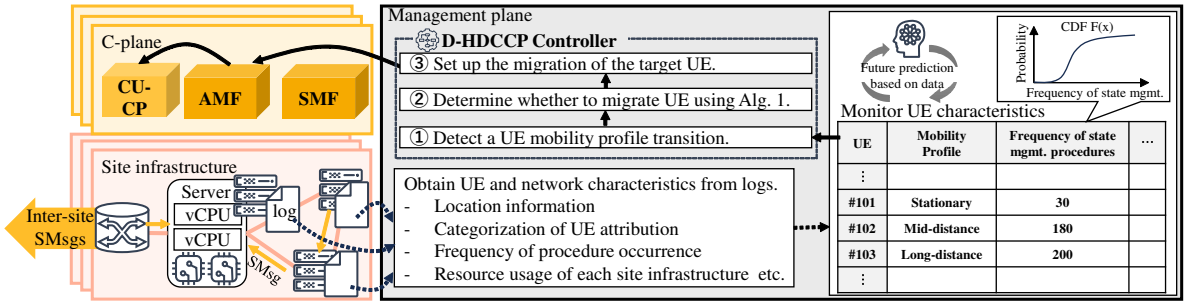


Fig. 4: System model of the D-HDCCP framework aligned with the 3GPP specification [13], including migration steps ① – ③.

management plane also predicts mobility profiles and future destinations over time, leveraging results from [23], to enable a responsive adaptation to transitions in these profiles.

The D-HDCCP controller in Fig. 4, functioning as a management plane component, performs green operations in the C-plane. Specifically, when a potential mobility profile change is detected (Step ① in Fig. 4), the controller evaluates the necessity of migration by analyzing UE characteristics and the target site infrastructure usage using Algorithm 1 (Step ②). Based on this assessment, the controller initiates the migration procedure illustrated in Fig. 5 (Step ③). Furthermore, the controller dynamically scales the C-plane NFIs at each site in response to the UE mobility distribution, following the stateful NFI scaling approaches as discussed in [24].

B. Objective Formulation

1) Network Model: Let $G(\mathcal{V}, \mathcal{E})$ denote an undirected graph representing site infrastructures, where $\mathcal{V}$ is the set of sites, and $\mathcal{E}$ is the set of backhaul links connecting two distinct sites $v, v' \in \mathcal{V}$ with $v \neq v'$. Similarly, let $H(\mathcal{N}, \mathcal{A})$ represent a 5G mobile network consisting of RAN and MCN NFIs over G, where $\mathcal{N}$ is the set of NFIs, and $\mathcal{A}$ is the set of virtual links connecting two distinct NFIs $n, n' \in \mathcal{N}$ with $n \neq n'$.

Let $\mathcal{R}_u$ denote the procedure sets requested by UE $u \in \mathcal{U}$. Each SMsg required to complete procedure $r \in \mathcal{R}_u$ is represented as an interaction between two distinct NFIs. Procedure r can be expressed as a sequence of k_c virtual links, that is, $\{(n_0, n_1), (n_1, n_2), \ldots, (n_{k_c-1}, n_{k_c})\}$. Thus, r is modeled as subgraph $r(\mathcal{N}_r, \mathcal{A}_r)$ of H over G, where $\mathcal{A}_r$ is the set of NFI connections, and $\mathcal{N}_r$ is the set of NFIs associated with $\mathcal{A}_r$. The NFs involved in each SMsg are pre-determined by 3GPP [9], [10], and the NFI selection follows specific rules: if the NFI holds the state of u, it is selected; otherwise, the NFI is selected based on policies such as load balancing.

2) D-HDCCP Objective: The objective is to minimize total power consumption P_{total} in the C-plane by UEs:

$$\text{Minimize} \quad P_{\text{total}} = \sum_{m} \sum_{u \in \mathcal{U}} \sum_{r \in \mathcal{R}_u^m} P_r \qquad (1)$$
$$\text{subject to} \quad \text{constraints (3) - (5)},$$

where $\mathcal{R}_u^m$ and P_r denote the procedure set during period $[mT, (m+1)T)$ and the power consumption by procedure r. Moreover, P_r is expressed in terms of $r = r(\mathcal{N}_r, \mathcal{A}_r)$ as:

$$P_r := \sum_{n \in \mathcal{N}_r} P_n + \sum_{a \in \mathcal{A}_r} P_a, \qquad (2)$$

where P_n and P_a represent the increase in power consumption in NFI n and its connection a, respectively. Note that $P_a = 0$ when a is not an inter-site SMsg, i.e, not a flew over $e \in \mathcal{E}$. Regardless of objective (1), the following constraints must be satisfied during every period $[mT, (m+1)T)$ for $\forall m \in \mathbb{N}$:

$$\mathcal{N}_{\text{CP}} := \{\mathcal{N}_{\text{CU-CP}}, \mathcal{N}_{\text{AMF}}, \mathcal{N}_{\text{SMF}}\} \subset \mathcal{N}.$$

$$\sum_{v \in \mathcal{V}} \xi_{n,v}^m = 1, \quad \forall n \in \bigcup_{r \in \mathcal{R}_u^m \setminus \{r_*\}} \mathcal{N}_r \cap \mathcal{N}_{\text{CP}}. \qquad (3)$$

$$\sum_{n \in \mathcal{N}_r \cap \mathcal{N}_{\text{CP}}} \xi_{n,v}^m = 3, \quad \forall r \in \mathcal{R}_u^m \setminus \{r_*\}, \ \forall v \in \mathcal{V}. \qquad (4)$$

$$\sum_{u \in \mathcal{U}} \gamma_{u,v}^m \leq C_v, \quad \forall v \in \mathcal{V}. \qquad (5)$$

Let $\mathcal{N}_{\text{CP}}$ be all CU-CP, AMF, and SMF instances in $\mathcal{N}$. Binary decision variable $\{0,1\}$ $\xi_{n,v}^m$ and $\gamma_{u,v}^m$ indicate if NFI $n \in \mathcal{N}$ is deployed on site $v \in \mathcal{V}$ and if UE $u \in \mathcal{U}$ is managed by C-plane site $v \in \mathcal{V}$, respectively. Constraint (3) ensures that for each procedure excluding migration procedure r_* of UE u, only one NFI of each NF type is utilized to preserve the state consistency of u. Constraint (4) enforces that procedures excluding migration procedure r_* are executed within a C-plane site to prevent inter-site SMsgs between C-plane NFIs. Constraint (5) restricts the number of UEs managed at each C-plane site with limited computing resources to guarantee stable SMsg processing. C_v denotes the maximum number of UEs that can be efficiently managed by C-plane site v.

This problem can be regarded as a variant of the Online Generalized Assignment Problem (OGAP) [25], a well-known NP-hard problem. OGAP involves dynamically assigning tasks to agents in an online setting to optimize a specific objective while satisfying constraints. In the D-HDCCP, tasks correspond to procedures requested by UEs, which occur randomly over time, while agents represent the C-plane sites with resource constraints. Thus, a greedy online allocation algorithm is developed, as finding an optimal C-plane site for processing a UE's procedures in real time becomes increasingly intractable with the scaling number of sites and UEs.

C. Adaptive Migration Strategy

The migration decision mechanism focuses on transitions between mobility profiles, specifically from mid-/long-moving to stationary. When a UE's mobility profile shifts from stationary to moving, migration to central C-plane site should be performed to facilitate not Inter AMF Handover but Intra CU Handover for each handover. This approach is

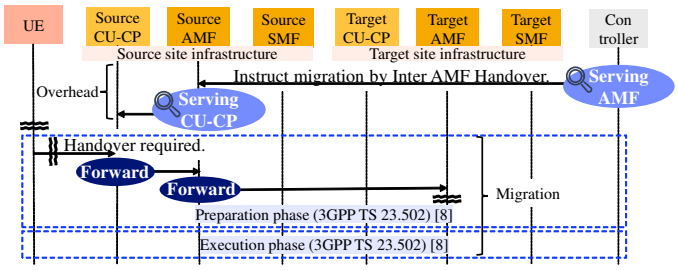


Fig. 5: Migration synchronized with handover (Step ③ in Fig. 4).

TABLE II: Comparison of total and inter-site SMsgs between independent and synchronized migrations.

	Mid-moving to stationary		Long-moving to stationary	
Migration	Independent	Synchronized	Independent	Synchronized
# of SMsgs	81 (—)	53 (34.6% ↓)	81 (—)	53 (34.6% ↓)
# of inter-site SMsgs	27 (—)	17 (37.0% ↓)	47 (—)	27 (42.6% ↓)

further justified by the fact that regional and central sites, which manage moving UEs, generally do not face computing resource constraints, i.e., $C_v \approx \infty$. On the other hand, when the mobility profile shifts from moving to stationary, migration to edge C-plane site should be avoided if the stationary period is short, since the reduction of sufficient inter-site SMsgs is not expected. The limited computing resources at edge sites further emphasize the need for a cautious assessment to ensure effective utilization. Therefore, we develop an assessment mechanism to implement this concept, along with an SMsg-efficient migration procedure.

1) Migration Procedure: To satisfy constraints (3) and (4), UE states associated with CU-CP, AMF, and SMF instances must be transferred to the corresponding instances at the target C-plane site in each migration. Consequently, this reallocation generates SMsgs comparable to those in Inter AMF Handover with these NFI switches. To mitigate this SMsg overhead, a migration method synchronized with the handover procedure is developed, as shown in Fig. 5.

Let $\hat{\mathcal{R}}_u \leftarrow \bigcup_{m=i}^{j} \mathcal{R}_u^m$, where $i < j$, be the set of procedures expected from $u \in \mathcal{U}$ during period $[iT, jT)$ from the time the mobility profile shifts from moving to stationary until it shifts back to moving. When migration procedures r_* and r'_* are executed independently, $\hat{\mathcal{R}}_u^* \leftarrow \hat{\mathcal{R}}_u \cup \{r_*, r'_*\}$, causing two Inter AMF Handover-equivalent procedures.

To minimize these SMsgs, migration methods synchronized with handover procedures are proposed. Specifically, handover procedures r_{ho} and r'_{ho} in $\hat{\mathcal{R}}_u$, which occur before and after mobility profile changes, are mapped to migration procedures utilizing Inter AMF Handover. Thus, $\hat{\mathcal{R}}_u^*$ is redefined as:

$$\hat{\mathcal{R}}_u^* \leftarrow \hat{\mathcal{R}}_u \setminus \{r_{ho}, r'_{ho}\} \cup \{r_*, r'_*\}, \tag{6}$$

where r_* and r'_* represent migration procedures. These procedures introduce a small overhead for controller setup instructions but, significantly reduce overhead by synchronizing migration with handover procedures. Table II is the substantial reduction effect in the case that r_{ho} is Intra AMF Handover.

The proposed migration procedure is implemented within the 3GPP framework by newly defining SMsgs from Application Function (AF in Fig. 1) to AMF and from AMF to CU-CP. Within the SBA, the controller, functioning as an AF [11], instructs the AMF instance managing a UE to execute Inter AMF Handover [10]. Since the AMF instance holds its state using a unique identifier, i.e., Subscription Permanent Identifier (SUPI), the controller can identify the serving AMF instance [12] and issue the handover instructions. Then, the AMF instance is configured to initiate Inter AMF Handover to the target C-plane site, after which the instruction is forwarded to the CU-CP instance. This process is feasible because the AMF instance holds the unique identifier linking the AMF and CU-CP instances for a UE, i.e., UE NGAP ID [9]. A similar configuration is applied to the CU-CP instance. Once these configurations are set, "Handover Required" SMsg is forwarded to the CU-CP and AMF instances, even if other handover options are available. The preparation and execution phases follow the same procedure as Inter AMF Handover with an SMF instance change.

To our knowledge, the concept of intentionally changing handover options to achieve a specific goal is novel.

2) Migration Threshold: Here, we discuss the metric used for migration criteria. After detecting changes in the UE mobility profile during $[iT, jT)$, the controller decides whether to initiate migration. This decision is based on a comparison of the power consumption between two procedure sets during $[iT, jT)$: $\hat{\mathcal{R}}_u^*$ (with migration) and $\hat{\mathcal{R}}_u$ (without migration). Thus, migration is triggered under the following conditions:

$$\sum_{r \in \hat{\mathcal{R}}_u^*} P_r - \sum_{r \in \hat{\mathcal{R}}_u} P_r < 0. \tag{7}$$

The difference between $\hat{\mathcal{R}}_u^*$ and $\hat{\mathcal{R}}_u$ lies in the reduction of inter-site SMsgs by procedures other than handovers and the replacement of handover procedures $\{r_{ho}, r'_{ho}\}$ with migration procedures $\{r_*, r'_*\}$. Since site infrastructures have equivalent hardware, the power consumption for processing SMsgs remains the same across both procedure sets, i.e., the difference in each P_n is 0. Thus, Eq. (7) is expanded using P_a as:

$$\begin{aligned} P_{r_*} + P_{r'_*} + \sum_{r(\mathcal{N}_r, \mathcal{A}_r) \in \hat{\mathcal{R}}_u^* \setminus \{r_*, r'_*\}} \sum_{a \in \mathcal{A}_r} P_a - P_{r_{ho}} \\ - P_{r'_{ho}} - \sum_{r(\mathcal{N}_r, \mathcal{A}_r) \in \hat{\mathcal{R}}_u \setminus \{r_{ho}, r'_{ho}\}} \sum_{a \in \mathcal{A}_r} P_a < 0. \end{aligned} \tag{8}$$

As described in Secs. II-A and II-C, since Service Request and AN Release occur frequently and alternately, and Registration and Deregistration almost never occur, migration threshold θ is expressed as:

$$\theta \geq \left\lceil \frac{P_{r_*} + P_{r'_*} - P_{r_{ho}} - P_{r'_{ho}}}{\sum_{r(\mathcal{N}_r, \mathcal{A}_r) \in \{r_{sr}, r_{ar}\} \subseteq \hat{\mathcal{R}}_u} \sum_{a \in \mathcal{A}_r} P_a} \right\rceil, \tag{9}$$

where θ is the number of occurrence of a set of Service Request r_{sr} and AN Release r_{ar} in $\hat{\mathcal{R}}_u$. Threshold θ highlights the frequency of idle/active state management procedures and the UE's mobility profile before migration, as inter-site SMsg volume directly affects power consumption

across backhaul links. Consequently, when state management procedure frequencies are comparable, migrating long-moving UEs proves more effective than migrating mid-moving UEs.

3) Online Migration Decision Algorithm: When computing resources at edge C-plane sites are constrained—particularly during the introduction or growth phase of edge computing deployment—it becomes impractical to accommodate all UEs to reallocate. Under these conditions, applying the migration threshold derived in Eq. (9) uniformly across all UEs leads to inefficiencies, potentially excluding those that should be prioritized for energy-efficient migration. To solve this issue while satisfying constraint (5), a classical statistical approach is used to dynamically adjust the migration threshold based on the distribution of state management frequencies, a key criterion for migration. The details are outlined in Algorithm 1.

Algorithm 1 is an online migration decision mechanism that determines whether to migrate UE processing to an edge or regional C-plane site v when the mobility profile of UE u shifts from moving to stationary. The candidate C-plane sites for migration, denoted as $\mathcal{V}_{can} \subset \mathcal{V}$, are selected based on each UE's mobility profile, with long-moving UEs having two candidate sites (lines 1–4). Next, the capacity constraint at v within $[iT, jT)$ is verified. If the constraint is met, the migration threshold is set according to Eq. 9 (lines 6–7). For mid-moving UEs, threshold θ is doubled due to their lower inter-site SMsg generation per state management procedure set. Specifically, mid-moving UEs produce 7 inter-site SMsgs per set of Service Request and AN Release procedures, whereas long-moving UEs generate 14 (line 10). The threshold is then refined using the inverse CDF F_v^{-1}, which models the distribution of state management procedures at antenna sites managed by v. The stationary UE-to-capacity ratio $|U^{t^*}_{v,\text{sta}}|/C_v$ is substituted into F_v^{-1} to capture the top $\left(1 - \frac{C_v}{|U^{t^*}_{v,\text{sta}}|}\right) \times 100\%$ of occurrences. The migration threshold is finally determined as the maximum of the value computed in line 10 and the CDF-based adjustment, ensuring compliance with Eq. 9 (line 11). If the number of state management procedures in $\hat{\mathcal{R}}_u$ exceeds this threshold, $\hat{\mathcal{R}}_u$ is updated to $\hat{\mathcal{R}}^*_u$ (lines 12–15).

Algorithm 1: Migration decision mechanism when mobility profile of UE u shifts from moving to stationary.

Input : Expected procedure set $\hat{\mathcal{R}}_u$ for $[iT, jT)$; Source C-plane site v';
Output: Procedure set $\hat{\mathcal{R}}^*_u$ (if migrated) or $\hat{\mathcal{R}}_u$ (if not); Target C-plane site v;

1 **if** u is mid-moving **then**
2 Set $\mathcal{V}_{can}$ as edge C-plane site managing the antenna site where u remains stationary;
3 **else if** u is long-moving **then**
4 Set $\mathcal{V}_{can}$ as edge and regional C-plane sites managing the antenna site where u remains stationary;
5 **for** $v \in \mathcal{V}_{can}$ **do**
6 **if** $\bigwedge_{t=i}^{j} \mathbf{1}\left(\sum_{u \in \mathcal{U}} \gamma^t_{u,v} < C_v\right) = 1$ **then**
7 Set threshold θ from Eq. (9); ▷ Sec. III-C2
8 $t^* \leftarrow \arg\max_{t \in \{i,\ldots,j-1\}} |U^t_{v,\text{sta}}|$;
9 **if** $C_v < |U^{t^*}_{v,sta}|$ **then**
10 $\theta \leftarrow (1 + \mathbf{1}(u \text{ is mid-moving})) \times \theta$;
11 $\theta \leftarrow \max\left(\theta, F_v^{-1}\left(1 - \frac{C_v}{|U^{t^*}_{v,\text{sta}}|}\right)\right)$;
12 **if** $|\hat{\mathcal{R}}_u \setminus \{r_{ho}, r'_{ho}\}| > 2 \times \theta$ **then**
13 $\hat{\mathcal{R}}^*_u \leftarrow \hat{\mathcal{R}}_u \setminus \{r_{ho}, r'_{ho}\} \cup \{r_*, r'_*\}$;▷ Sec. III-C1
14 $\gamma^t_{u,v'} \leftarrow 0,\ \forall t \in [i,j)$; $\gamma^t_{u,v} \leftarrow 1,\ \forall t \in [i,j)$;
15 **return** $\hat{\mathcal{R}}^*_u$, v;
16 **return** $\hat{\mathcal{R}}_u$, v';

D. Use-case Discussion

Commuter UEs serve as an ideal use case for D-HDCCP, which relies on mobility prediction, as their structured daily mobility patterns, characterized by movement and stationary phases, enhance mobility prediction effectiveness, as stated in [23]. While the prediction model in [23] identifies the next cell site, D-HDCCP requires the next antenna site (a set of cell sites), thereby relaxing prediction constraints. Furthermore, mobility analysis in the Tokyo Metropolitan Area (TMA) [19] reports that commuter UEs remain stationary for approximately eight hours after commuting. This prolonged stationary period facilitates D-HDCCP's efficient management of commuter UEs, as it potentially allows for a sufficient number of state management procedures to benefit from migration.

We believe that D-HDCCP can also be applied to IoT devices. In the use-cases listed in [1], autonomous mobile robots and vehicles that follow pre-programmed navigation paths and have periodic stationary intervals, as well as wearable devices and bio-nano things used by individuals with commuter-like mobility patterns, may present opportunities for migration.

IV. Simulation-based Analysis

To evaluate the energy-saving potential of D-HDCCP, we focus on commuter UEs, a use case discussed in Sec. III-D. Statistical analyses of commuters in [18], [19] also support this choice. We first show that D-HDCCP outperforms LCP and HDCCP in migration necessity when Service Request and AN Release are sufficiently observed. Then, extensive simulations based on real-world data demonstrate D-HDCCP's C-plane energy savings in both fully deployed and early-stage edge computing scenarios. Finally, we analyze the migration decision mechanism and offer operational insights for threshold policies. For evaluation, a discrete-event simulator replicating UEs and mobile networks is implemented in Python.

A. Simulation Setup

To evaluate P_n and P_a during period $[mT, (m+1)T)$ in Eq. (2), we adopt widely-used linear power consumption models, as referenced in [26] and [4]. Since the CPU is the primary source of power consumption, the increase in power consumption to process an SMsg is represented below:

$$P_n = \frac{z_n}{T} \times w_{\text{CPU}},$$

TABLE III: Simulation parameters.

Parameter	Value	Parameter	Value
w_{CPU}	3.33×10^{-8}	$z_n, \forall n \in \mathcal{N}_{\text{Other}}$	1.61×10^5
w_{BH}	4.00×10^{-6}	$l_{(n,n')}, \forall n' \in \mathcal{N}_{\text{CU-CP}}$	176
$z_n, \forall n \in \mathcal{N}_{\text{CU-CP}}$	5.20×10^4	$l_{(n,n')}, \forall n' \in \mathcal{N}_{\text{AMF}}$	149
$z_n, \forall n \in \mathcal{N}_{\text{AMF}}$	1.72×10^5	$l_{(n,n')}, \forall n' \in \mathcal{N}_{\text{SMF}}$	157
$z_n, \forall n \in \mathcal{N}_{\text{SMF}}$	1.49×10^5	$l_{(n,n')}, \forall n' \in \mathcal{N}_{\text{Other}}$	170

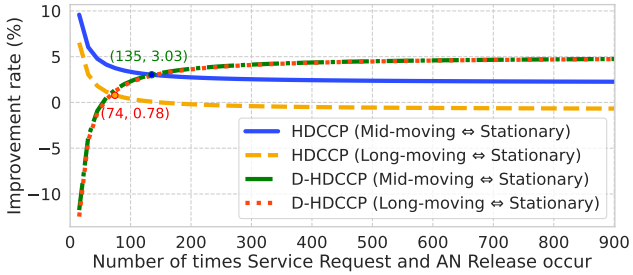


Fig. 6: Power consumption improvement rate of HDCCP (static allocation) and D-HDCCP (dynamic reallocation) to LCP.

where z_n is the number of CPU cycles required to process an SMsg in NFI n, and w_{CPU} is the power consumption per unit of CPU cycles. Their units are [Cycles] and [Watts per Cps].

$$P_a = \frac{l_a \times d(a)}{T} \times w_{\text{BH}}, \quad a = (n, n'),$$

where l_a is the packet length of an SMsg, w_{BH} is the power consumption per unit of inter-site backhaul link capacity. Their units are [Bytes] and [Watts per Bps]. $d(a)$ indicates the number of site traversals for the SMsg: when forwarded from the central site to a edge site (or vice versa), it passes through the regional site, so $d(a) = 2$; if forwarded within the same site, $d(a) = 0$; otherwise, $d(a) = 1$.

Table III lists the simulation parameters. w_{CPU} and w_{BH} are sourced from [26] and [4]. The values for z_n are measured using the Read Time Stamp Counter (RDTSC) [27], with UERANSIM [28] and free5GC [29] running on a server equipped with a 2.30 GHz Intel(R) Xeon(R) processor. The values for $l_{(n,n')}$ are obtained through the packet capture and show no significant deviation from the values reported in [30].

B. Comparison Analysis of LCP, HDCCP, and D-HDCCP

Figure 6 presents the improvement rate in C-plane power consumption for a single UE as it transitions between different mobility profiles—specifically, from long-/mid-moving to stationary and back to long-/mid-moving—under HDCCP and D-HDCCP, relative to LCP. In D-HDCCP, a UE is reallocated to an edge C-plane site when its mobility profile becomes stationary. This evaluates the impact of $\hat{\mathcal{R}}_u$ and $\hat{\mathcal{R}}_u^*$.

D-HDCCP demonstrates superior energy efficiency compared to HDCCP when the number of state management procedures (a combination of Service Request and AN Release) exceeds 135 for mid-moving UEs and 74 for long-moving UEs. These values are the migration thresholds for each UE mobility profile in scenarios where resource constraints are not considered. Furthermore, HDCCP becomes less efficient than LCP when long-moving UEs execute more than 150 state management procedures, underscoring the limitations of a static allocation strategy. These findings emphasize the necessity of adaptive migration strategies to accommodate

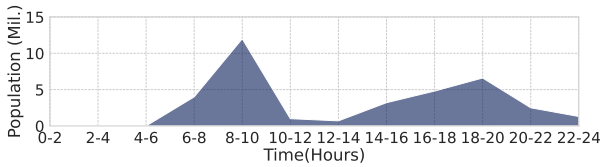
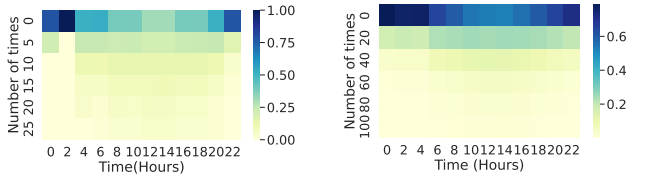


Fig. 7: Commuter UE movement period in the TMA, based on [19].



(a) Handover procedures. **(b)** Service Req. & AN Release.

Fig. 8: Heat maps showing the occurrence probabilities of each procedure, with data from [18] mapped to Poisson distributions.

variations in UE mobility profiles. The need for migration is particularly pronounced for long-moving UEs, where inefficient static allocations result in increased power consumption.

As the number of state management procedures increases, the improvement rate of D-HDCCP asymptotically converges to 5.0%. This stabilization occurs because the improvement rate over LCP for stationary UEs continuously managed by edge C-plane sites inherently reaches this limit.

While migration effectiveness is evident, it is crucial to assess the practicability of multiple state management procedures occurring while the UE remains stationary. As detailed in Sec. III-D and Fig. 7, commuter UEs typically remain stationary for about eight hours post-commute. According to [18], the average Service Request frequency during this time is about 20 per hour, totaling roughly 160 procedures. This value surpasses the migration thresholds for mid- and long-moving UEs, reinforcing commuter UEs as a strong use case for D-HDCCP.

C. Analysis in Large-scale Scenario with Resource Constraint

Scenario. The effectiveness of D-HDCCP is evaluated in the TMA, assuming a mobile network with 18 million commuter UEs [19], consisting of a single central site and multiple regional and edge sites. This single central site configuration meets the distance constraints outlined in [3]. A UE's mobility behavior follows the pattern in Fig. 7, with movement start and stop times modeled accordingly. After commuting, UEs remain at one of $K = 174$ antenna sites in the urban area, controlled by edge C-plane sites, before returning to their respective residential areas, which are distributed across various antenna sites. This setup effectively models commuter movement from residential areas to central workplaces, as characterized in [19]. Three scenarios are examined under following conditions: (A) introduction phase – 20% deployment to edge sites ($C_v = \lfloor 18 \text{ million} \div K \times 0.2 \rfloor$), (B) growth phase – 70% deployment ($C_v = \lfloor 18 \text{ million} \div K \times 0.7 \rfloor$), and (C) full deployment – Complete deployment to all edge sites. The value of $K = 174$ corresponds to the estimated number of edge sites in Tokyo, Japan's capital, based on data from [31].

The procedures requested by UEs are randomly determined based on Fig. 8. The number of handover procedures per UE,

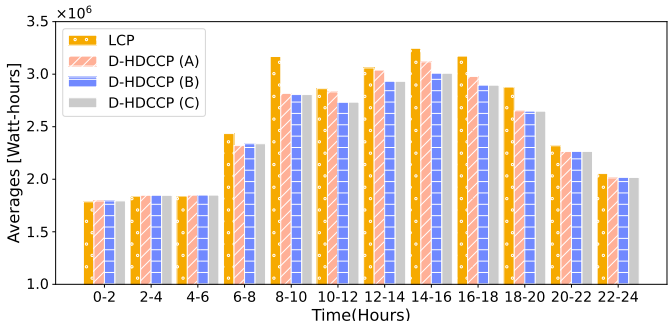


Fig. 9: C-plane power consumption of commuter UEs in the TMA across different deployment phases to edge sites.

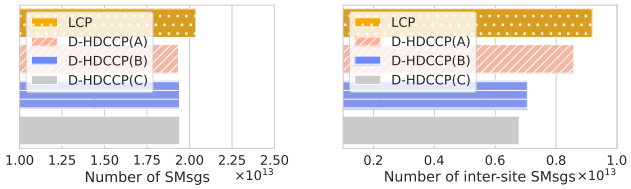
TABLE IV: Daily power consumption (in $\times 10^6$ Watt-hours).

LCP	D-HDCCP (A)	D-HDCCP (B)	D-HDCCP (C)
30.7 (—)	29.5 (3.9% ↓)	29.1 (5.1% ↓)	29.0 (5.2% ↓)

k_{ho}, is derived from Fig. 8(a). Each UE follows a predefined handover sequence, performing Intra CU Handover twice and Intra AMF Handover once in a loop, resulting in a total of k_{ho} handovers. This setup simulates a scenario where, under the LCP configuration, a UE is definitively reallocated to a different regional C-plane site after undergoing Intra CU Handover twice. Thus, in this handover configuration, UEs performing Intra CU Handover once or twice are classified as mid-moving, while UEs with more than two handovers are classified as long-moving. In both stationary and moving profiles, the number of Service Request and AN Release procedures is determined according to Fig. 8(b). Additionally, CDF $F_v(x)$, utilized in Algorithm 1, is derived from this distribution. Since the UE's destination is randomly determined, we can effectively reproduce the degree of uncertainty in the prediction by unifying $F_v(x)$ at any edge site $v \in \mathcal{V}$.

Results. Figure 9 and Table IV present the average C-plane power consumption of commuter UEs at two-hour intervals and their cumulative sum representing daily power consumption, based on ten different random seeds. Results show that D-HDCCP with full deployment consistently achieves lower power consumption compared to LCP, with a peak reduction of 11.0% during the busiest travel period (8–10 AM) and an overall reduction of 5.2%. This efficiency is realized by reducing total SMsgs by 4.7% and inter-site SMsgs by 26.1%, as shown in Figs. 10(a) and 10(b).

We conduct a detailed analysis for each time period. From 0–6 AM, all scenarios exhibit equivalent power consumption since migrations provide no benefit during this period, as UEs are not activated and do not generate state management procedures after returning home, as shown in Fig. 8(b). During 6–10 AM and 4–8 PM, D-HDCCP outperforms LCP by effectively reducing SMsgs by facilitating Intra CU Handover. Between 10 AM and 6 PM, scenarios (B) and (C) achieve higher energy savings than LCP and (A). This is because many state management procedures are processed at edge C-plane sites through migration and significant inter-site SMsgs reduction is realized, as shown in Fig. 10(b). The similarity in power consumption between scenarios (B) and (C) indicates that migration effectively accommodates nearly all UEs that



(a) Total SMsgs. **(b)** Inter-site SMsgs.

Fig. 10: Daily total and inter-site SMsg counts.

TABLE V: F1 score in each migration decision policy.

	(A) Introduction	(B) Growth	(C) Full deployment
FIFO	Recall: 0.183 Precision: 0.162 F1 Score: 0.172	Recall: 0.733 Precision: 0.706 F1 Score: 0.719	Recall: 1.0 Precision: 0.737 F1 Score: 0.848
Static threshold	Recall: 0.226 Precision: 0.202 F1 Score: 0.213	Recall: 0.920 Precision: 0.896 F1 Score: 0.908	Recall: 1.0 Precision: 1.0 F1 Score: 1.0
Adjusted threshold	Recall: 0.829 Precision: 0.901 F1 Score: 0.863	Recall: 0.945 Precision: 0.977 F1 Score: 0.961	Recall: 1.0 Precision: 1.0 F1 Score: 1.0

can benefit from it. These findings suggest that a sufficient deployment of edge C-plane sites in urban areas is crucial for green operations by D-HDCCP, even without full deployment.

D. Analysis of Migration Decision Mechanism

The effectiveness of Algorithm 1 is evaluated under the same simulation scenario as Sec. IV-C. Three migration decision policies are compared: "FIFO policy", which omits lines 7 to 12 of Algorithm 1 and applies a first-in, first-out approach; "static threshold policy", which omits lines 8 to 11 and uses the threshold derived from Eq. 9; and "adjusted threshold policy", which fully implements Algorithm 1, incorporating adaptive threshold adjustments using CDF $F_v(x)$.

Table V presents the F1 score [32], which represents the harmonic mean of precision and recall. Here, precision is the ratio of correctly reallocated optimal UEs to the total that should be reallocated, while recall is the ratio of correctly reallocated optimal UEs to the total actually reallocated.

The evaluation shows that during the introduction phase, the adjusted threshold policy achieves the highest effectiveness, while other policies perform significantly worse. This highlights the importance of careful migration management when edge computing resources are limited. In the growth phase, the static threshold policy achieves comparable performance, as the reduction in inter-site SMsgs becomes marginal. This suggests that with expanded edge computing deployment, similar accuracy can be achieved even without strict threshold verification, indicating that a fixed threshold can also be effective under sufficient resource availability.

In the full deployment scenario, where all UEs eligible for migration can be successfully migrated, the F1 score reaches 1 for policies that incorporate a threshold-based decision mechanism. In contrast, the FIFO policy, which migrates all commuter UEs as soon as their mobility profile shifts to stationary, also results in migrations for UEs that do not meet the predefined threshold criteria. However, despite this inefficiency, the difference in daily power consumption between FIFO and other policies is minimal, with only a

0.02% downgrade. This minor difference is because many UEs are observed to be close to the migration threshold. These findings indicate that all UEs categorized as commuters can optimistically undergo migration without significant concerns regarding energy efficiency degradation.

V. Conclusion

We proposed Dynamic Hybrid Distributed and Centralized C-plane Placement (D-HDCCP) within the 3GPP framework to enable green C-plane operations. D-HDCCP dynamically allocates UEs to geo-distributed C-plane sites based on their mobility profiles and determines whether migration for re-allocation should be performed as mobility transitions. By reducing total and inter-site SMsgs, D-HDCCP effectively lowers power consumption, which addresses a key challenge in 5G/6G mobile networks. Additionally, commuter UEs were identified as a representative use case of D-HDCCP, demonstrating its effectiveness and operational insights across different deployment phases in edge computing environments.

In future work, we plan to develop a testbed to evaluate the benefits and challenges of D-HDCCP in actual conditions. This will further validate its feasibility in terms of metrics such as prediction accuracy and real-time performance.

Appendix: Mobility Profiles

To efficiently reduce inter-site SMsgs for a UE, mobility profiles are classified in accordance with the hierarchical site structure. Specifically, these profiles are designed based on a k_m-level hierarchical site architecture, where level-1 sites represent central sites, and level-k_m sites correspond to far-edge sites. Consequently, UE set $\mathcal{U}$ is divided into k_m subsets: $\{\mathcal{U}_1, \mathcal{U}_2, \ldots, \mathcal{U}_{k_m}\}$, where $\mathcal{U}_i \cap \mathcal{U}_j = \emptyset$ $(i \neq j)$. This division is based on whether a UE moves within antenna sites controlled by the level-i C-plane site, i.e., the UE does not cross into antenna sites controlled by higher-level ($> i$) C-plane sites. The aim is to reduce SMsgs associated with handovers while minimizing inter-site SMsgs. Stationary UEs in U_{k_m} pursue only to minimize inter-site SMsgs for state management.

This allocation rule greatly improves upon the approach proposed in [8], which classifies mobility profiles solely as either high or low based on handover frequency. By implementing it, cases where UE u is erroneously allocated to a level-1 C-plane site despite being better suited for U_i with $i > 1$ are prevented. This prevention reduces $\frac{k_m - 1}{k_m - i}$-fold inter-site SMsgs.

References

[1] M. Banafaa, I. Shayea, J. Din, M. H. Azmi, A. Alashbi, Y. I. Daradkeh *et al.*, "6g mobile communication technology: Requirements, targets, applications, challenges, advantages, and opportunities," *Alex. Eng. J.*, vol. 64, pp. 245–274, 2023.

[2] M. Q. Khan, "Signaling storm problems in 3gpp mobile broadband networks, causes and possible solutions: A review," in *IEEE iCCECE*, 2018, pp. 183–188.

[3] S. Mondal and M. Ruffini, "Optical front/mid-haul with open access-edge server deployment framework for sliced o-ran," *IEEE Trans. Netw. Serv. Manag.*, vol. 19, no. 3, pp. 3202–3219, 2022.

[4] J. Zhang, X. Zhang, and W. Wang, "Cache-enabled software defined heterogeneous networks for green and flexible 5g networks," *IEEE Access*, vol. 4, pp. 3591–3604, 2016.

[5] A. Tabiban, H. A. Alameddine, M. A. Salahuddin, and R. Boutaba, "Signaling storm in o-ran: Challenges and research opportunities," *IEEE Commun. Mag.*, vol. 62, no. 6, pp. 58–64, 2024.

[6] J. Jeong, D. Roeland, J. Derehag, Å. A. Johansson, V. Umaashankar, G. Sun *et al.*, "Mobility prediction for 5g core networks," *IEEE Commun. Stand. Mag.*, vol. 5, no. 1, pp. 56–61, 2021.

[7] X. Yan, M. Ma, and R. Su, "Efficient group handover authentication for secure 5g-based communications in platoons," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 3, pp. 3104–3116, 2022.

[8] M. Kurata, A. Ikami, S. Itahara, and M. Suzuki, "Signaling storm mitigation by geographically distributed c-plane nf placement and routing," in *IEEE NetSoft*, 2024, pp. 100–108.

[9] 3GPP, "NG-RAN; Architecture description," TS 38.401.

[10] ——, "Procedures for the 5G System (5GS)," TS 23.502.

[11] ——, "System architecture for the 5G System (5GS)," TS 23.501.

[12] ——, "Architecture enhancements for 5G System (5GS) to support network data analytics services," TS 23.288.

[13] ——, "Telecommunication management; Management concept, architecture and requirements for mobile networks that include virtualized network functions," TS 28.500.

[14] V. Jain, H.-T. Chu, S. Qi, C.-A. Lee, H.-C. Chang, C.-Y. Hsieh *et al.*, "L25gc: A low latency 5g core network based on high-performance nfv platforms," in *ACM SIGCOMM*, 2022, pp. 143–157.

[15] M. Corici, E. Troudt, and T. Magedanz, "An organic 6g core network architecture," in *ICIN*, 2022, pp. 1–7.

[16] E. Goshi, R. Stahl, H. Harkous, M. He, R. Pries, and W. Kellerer, "Pp5gs—an efficient procedure-based and stateless architecture for next-generation core networks," *IEEE Trans. Netw. Serv. Manag.*, vol. 20, no. 3, pp. 3318–3333, 2023.

[17] M. A. Kirchberger and L. Pohl, "Technology commercialization: a literature review of success factors and antecedents across different contexts," *J. Technol. Transf.*, vol. 41, pp. 1077–1112, 2016.

[18] J. Meng, J. Huang, Y. C. Hu, Y. Koral, X. Lin, M. Shahbaz *et al.*, "Characterizing and modeling control-plane traffic for mobile core network," *arXiv preprint arXiv:2212.13248*, 2022.

[19] K. Liu, Y. Murayama, and T. Ichinose, "A multi-view of the daily urban rhythms of human mobility in the tokyo metropolitan area," *J. Transp. Geogr.*, vol. 91, p. 102985, 2021.

[20] S. Kianpisheh and T. Taleb, "A survey on in-network computing: Programmable data plane and technology specific applications," *IEEE Commun. Surv. Tutor.*, vol. 25, no. 1, pp. 701–761, 2022.

[21] P. Parastar, A. Lutu, Ö. Alay, G. Caso, and D. Perino, "Spotlight on 5g: Performance, device evolution and challenges from a mobile operator perspective," in *IEEE INFOCOM*, 2023, pp. 1–10.

[22] M. Suzuki, T. Kitahara, S. Ano, and M. Tsuru, "Group mobility detection and user connectivity models for evaluation of mobile network functions," *IEEE Trans. Netw. Serv. Manag.*, vol. 15, no. 1, pp. 127–141, 2017.

[23] M. Yan, S. Li, C. A. Chan, Y. Shen, and Y. Yu, "Mobility prediction using a weighted markov model based on mobile user classification," *Sensors*, vol. 21, no. 5, 2021.

[24] S. Woo, J. Sherry, S. Han, S. Moon, S. Ratnasamy, and S. Shenker, "Elastic scaling of stateful network functions," in *15th USENIX Symposium on NSDI*, 2018, pp. 299–312.

[25] H. Liu, H. Zhang, K. Luo, Y. Xu, Y. Xu, and W. Tong, "Online generalized assignment problem with historical information," *Computers & Operations Research*, vol. 149, p. 106047, 2023.

[26] A. Gandhi, M. Harchol-Balter, R. Das, and C. Lefurgy, "Optimal power allocation in server farms," *ACM SIGMETRICS Perform. Eval. Rev.*, vol. 37, no. 1, pp. 157–168, 2009.

[27] Y. Yang, H. Jiang, Y. Wu, Y. Lv, X. Li, and G. Xie, "C2qos: Cpu-cycle based network qos strategy in vswitch of public cloud," in *IFIP/IEEE IM Symposium*, 2021, pp. 438–444.

[28] "Ueransim," https://github.com/aligungr/UERANSIM, accessed: 2025-02-07.

[29] "free5gc," https://www.free5gc.org, accessed: 2025-02-07.

[30] D. M. Manias, A. Chouman, and A. Shami, "An nwdaf approach to 5g core network signaling traffic: Analysis and characterization," in *IEEE GLOBECOM*, 2022, pp. 6001–6006.

[31] NTT East Corp., "Dsl service area information," https://www.ntt-east.co.jp/info-st/info_dsl/area.html, accessed: 2025-02-07.

[32] Y. Sasaki, "The truth of the f-measure, manchester: Mib-school of computer science," *University of Manchester*, vol. 25, 2007.