

ARTICLE TYPE

CU Placement over a Reconfigurable Wireless Fronthaul in 5G Networks with Functional Splits[†]

Davit Harutyunyan, Roberto Riggio¹ | Slawomir Kuklinski² | Toufik Ahmed³

¹FBK CREATE-NET, Trento, Italy

²Orange Polska, Warsaw, Poland

³CNRS-LaBRI, University of Bordeaux, Bordeaux, France

Correspondence

*Davit Harutyunyan, FBK CREATE-NET, Trento, Italy. Email: d.harutyunyan@fbk.eu

Summary

Mobile network operators are currently facing a tremendous increase in the level of data traffic. Although cell size reduction is one of the most common ways used to accommodate such traffic demand, densely deployed small cells also dramatically increase the level of inter-cell interference. By centralizing baseband signal processing at powerful computing infrastructures, called Centralized Unit (CU) pools, Cloud Radio Access Network (C-RAN) enables advanced coordination algorithms to be employed in dense small cell networks. In C-RAN, due to stringent bandwidth and latency requirements at the fronthaul links, the optical fiber, thanks to its bandwidth and latency characteristics, continues to be the most prevalent fronthaul medium option. Nevertheless, the optical fiber is one of the fronthaul options, while C-RAN (PHY-RF split) is one of the functional splits that can be defined each coming with different fronthaul requirements.

In this paper, we formulate and solve a dynamic CU placement problem for mobile networks as an Integer Linear Programming (ILP) problem. In the considered network, CU pools are placed at the edges of the network, and a reconfigurable Millimeter Wave (MMW) wireless fronthaul links are used in order to provide Decentralized Units (DUs) with connectivity. We study the impact of different functional splits on the placement cost and on the acceptance ratio using different substrate networks. Lastly, we propose and evaluate a CU placement heuristic algorithm using a numerical simulator. The results reveal that the optimal functional split selection can lead to significant resource utilization benefits in the RAN.

KEYWORDS:

Mobile Networks, C-RAN, Functional Split, CU Placement, Wireless Fronthaul, MMW

1 | INTRODUCTION

Mobile data traffic has been snowballing over the recent years. Cisco's analysis shows that the global mobile data traffic grew 63% in 2016, ranging from 44% growth in the North America to 96% in the Middle East and Africa. According to its forecast, the global mobile data traffic is expected to increase from 7 exabytes in 2017 up to 49 exabytes in 2021¹, while Ericsson's forecast shows that by 2022 the global mobile data traffic will reach up to 71 exabytes out of which around 75% accounting for

[†]Research leading to these results received funding from the European Union's H2020 Research and Innovation Action under Grant Agreement H2020-ICT-761592 (5G-ESSENCE Project).

video traffic². In order to satisfy this formidable traffic demand, costly upgrades in Mobile Network Operators' (MNOs) network become inevitable.

In legacy mobile networks, two main components of base stations, the radio units and the baseband processing units, are usually deployed together. The goal of such a deployment is to reduce high signal losses associated with coaxial cables that are normally used to interconnect these components. In order to circumvent this co-location constraint, MNOs migrated to the Distributed Radio Access Network (D-RAN) architecture, in which coaxial cables are substituted by optical fibers, and a digital interface is used to carry the In-phase/Quadrature (I/Q) samples between the baseband unit and the radio unit. However, since in the D-RAN architecture a Long Term Evolution (LTE) base station, called eNB, cannot share its frequency resources with other eNBs. These scarce frequency resources will be underutilized most of the time in a day due to the spatiotemporally fluctuating traffic demand.

In order to tackle the aforementioned problem, the Cloud RAN (C-RAN)³ architecture has been recently proposed. C-RAN is capable of reducing the Total Cost of Ownership (TCO) of mobile networks while, at the same time, optimizing network capacity, coverage, and power consumption. Such goals are achieved in C-RAN by separating baseband units, termed Centralized Units (CUs) from the radio units, termed Distributed Units (DUs) and combining CUs in powerful computing infrastructures, termed CU pools¹. CU pools can run on a number of general-purpose-of-the-shelf servers deployed in centralized locations.

The separation of the signal processing functionalities in the RAN protocol stack between the CU pool and the DU is known as functional split⁵. Traditionally, in the C-RAN all baseband signal processing (i.e., the entire RAN protocol stack) is taking place at the CU pools, while the DUs are only in charge of basic Radio Frequency (RF) processing such as signal amplification, analog-to-digital (AD) and reverse conversion. This kind of functional split between the CU pool and the DU is called Physical layer RF (PHY-RF) split. The PHY-RF split can be employed to overcome the spikes in traffic demand by dynamically adding CUs when and where required. Moreover, the PHY-RF split can improve the quality of experience of users at the edges of the cells. The main drawback of PHY-RF split, however, lies in the tight bandwidth and latency requirements imposed on the fronthaul links (i.e., the links interconnecting CU pools with DUs) where protocols such as Common Public Radio Interface (CPRI)⁶ are typically used to carry the I/Q samples. For example, CPRI rate of ≈ 2.5 Gbps is required on the fronthaul links in order to carry signals over a 20 MHz LTE Frequency Division Duplex (FDD) channel using a 2x2 Multiple-Input and Multiple-Output (MIMO) antenna configuration. As a consequence, providing fiber-based CPRI fronthaul links for ultra-dense small cells cannot simply be a viable option for MNOs.

Nevertheless, PHY-RF split is one among the several functional splits that can be defined in the RAN protocol stack between CU pools and DUs. In general, the lower-layer is the functional split option within the RAN protocol stack, the more are the benefits of resource centralization, however, the stricter is the fronthaul requirement. For example, the PHY-RF split allows MNOs to employ advanced inter-cell coordination techniques such as Joint Transmission/Reception⁷, while the splits higher than the PHY layer cannot support the aforementioned techniques. Sec. 2 provides a detailed description of the functional splits considered in this work.

Regarding the fronthauling options, wireless fronthaul has gained popularity as a cost-efficient solution for delivering connectivity between small cells. Recent progress in the Millimeter Wave (MMW) communication operating at the E-band (70–80 GHz) enables a few Gbps of bandwidth to be carried up to one km of distance. This makes MMW communication an appropriate technology to be used for building the fronthaul networks of dense mobile network deployments. Besides, devices operating in the E-band can leverage on compact antennas, allowing to pack several interfaces in a small form factor.

In this paper, we formulate and solve a virtual network embedding problem, which is also called a CU placement problem in the C-RAN scenario. Virtual networks are requested by Mobile Virtual Network Operators (MVNOs) and embedded by MNOs, which are considered as Infrastructure Providers (InPs). In the InP's physical network, also called the substrate network, CU pools are placed at the edges of the network, possibly co-located with macro cells and/or distributed clouds, while a reconfigurable MMW fronthaul is used in order to provide DUs with fronthaul connectivity. The MMW fronthaul network leverages on steerable directional antennas in order to adapt its topology to different usage scenarios. The reconfigurability of the substrate network allows virtual network requests to be re-embedded in case of changing traffic patterns (e.g., daytime versus nighttime traffic) or when new requests arrive. This allows the traffic of low-utilized MMW fronthaul links to be aggregated, shutting down the unnecessary MMW interfaces and, therefore, reducing the power consumption at C-RAN.

In order to find the optimal solution, the CU placement problem is formulated and solved using an Integer Linear Programming (ILP) technique, while a heuristic algorithm is also proposed to address the scalability issue of the ILP-based algorithm. This

¹Notice that the 3GPP⁴ terminology with a slight modification is used throughout this article. Specifically, the term CU is used for a BaseBand Unit (BBU) and the term CU pool is used as a BBU pool.

paper extends our previous work⁸ by generalizing the problem formulation to the substrate/virtual networks with different functional splits and by extending the evaluation with a random substrate network topology. Moreover, in this work, the problem formulation and the heuristic are extended in order to allow the re-use of substrate links by multiple virtual network requests thus improving the overall network utilization.

The rest of this paper is structured as follows. The considered functional splits are introduced in Sec. 2. The related work is discussed in Sec. 3. The substrate network model and the virtual network request model are detailed in Sec. 4. The ILP problem and the heuristic are introduced in Sec. 5. The numerical results are reported in Sec. 6. Finally, Sec. 7 draws the conclusions pointing out future work.

2 | FUNCTIONAL SPLITS

The functional split problem has attracted significant attention from both the academia and the industry. There are in fact different approaches to small cell virtualization in terms of the point at which base stations operations are decomposed into physical and virtual. In this section, we introduce the LTE functional splits that are considered in this work and discuss their pros and cons. Figure 1 illustrates the basic signal processing functionalities of the LTE network stack in the uplink² direction, highlighting the points at which a functional split is possible, while Table 1 compares different functional splits in terms of fronthaul bandwidth and latency requirements³.

PHY–RF Split. The PHY–RF split corresponds to full resource centralization with all baseband processing taking place at the CU pool, leaving the RF functions (e.g., AD/DA conversion, and signal amplification) at the DU side. While this functional split provides several advantages in terms of energy efficiency, computational diversity, improved spectral efficiency³, its tight fronthaul requirements can undermine its economical convenience.

PHY Split. By placing some of the physical layer functionalities such as Fast Fourier Transform (FFT) / Inverse Fast Fourier Transform (IFFT), subcarrier mapping/demapping, signal equalization and MIMO processing at the DUs, it is possible to significantly relax the fronthaul requirements in terms of both bandwidth and latency. As it can be seen in Table 1, taking the requirements of the PHY–RF split as a baseline, the PHY split allows the fronthaul bandwidth requirements to be reduced by a factor of 2.5. This is due to the removal of the Cyclic Prefix (CP) from the baseband signal and due to the fact that only received signals of the allocated PRBs are forwarded to the CU pool, therefore, providing a statistical multiplexing gain. Similarly, the fronthaul latency requirements are also relaxed by a factor of 8 when the PHY split is used. Notice however how these requirements are relieved at the expense of reduced resource centralization gain. For example, compared to the PHY–RF split, Coordinated Multipoint (CoMP) features such as joint transmission/reception can no longer be employed with the PHY split⁷. This can result in lower performances especially for cell-edge users, which are the ones that benefit the most from the interference reduction/cancellation features enabled by CoMP.

MAC Split. In this case, the Hybrid Automatic Repeat Request (HARQ) procedure is taking place at the DU while the rest of the Medium Access Control (MAC) functions along with the upper layers are consolidated at the CU pool. Compared to the PHY–RF split, the MAC split allows relaxing the latency requirements by a factor of 24 and the bandwidth requirements by a factor of 16.5. Functions such as joint decoding can no longer be exploited, however, joint scheduling and joint path selection are still possible.

PDCP–RLC Split. In this case, all functionalities are taking place at the DUs with the only exception of the Packet Data Convergence Protocol (PDCP) layer which is centralized at the CU pool. PDCP – Radio Link Control (PDCP–RLC) split and MAC split have the same fronthaul bandwidth requirement, since only the RLC header information bits, which are negligible, are removed from the signal before transmitting it to the PDCP layer. However, these two splits differ significantly in terms of fronthaul latency requirement. The main advantages of PDCP–RLC split are the capability to enable WiFi offloading and load balancing.

Selecting an optimal functional split option is not a trivial task since various factors, such as energy efficiency, traffic demand, service latency constraint, have to be considered in order to select the actual functional split option. HARQ, for example, imposes a strict latency requirement. When a functional split is selected such that the scheduler resides at a CU pool then HARQ interleaving⁹ can be used to stretch the strict one-way latency requirement of HARQ from 4ms to 8ms. This is a side benefit which

²The functional splits considered in this work are symmetrical. Thus, all the considerations made for the uplink also apply to the downlink direction.

³The fronthaul bandwidth and latency requirements along with the latency classification for all the splits in the table are taken from SCF⁹.

entails a reduction of the peak data rate. Alternatively, the 4ms latency requirement of HARQ can be easily met by adopting another split in which the scheduler resides at DUs. However, this would result in reduced resource centralization benefits.

3 | RELATED WORK

The continuous increase in the mobile data traffic demand required costly upgrades in mobile networks in order to support the traffic demand. On the other hand, nowadays the frequency resources (i.e., LTE bands) of mobile networks are not used efficiently since those resources are allocated to eNBs with the goal of meeting the traffic demand at peak hours. As a result, due to fluctuating traffic demands, those resources might be underutilized for a long time. C-RAN is one of the technologies that is able to cope with the aforementioned challenges^{3,10}.

CU Placement. A sizable body of work has been published on the CU placement problem and on the C-RAN technology in recent years^{11,12,13,14,15}. Namba et al¹² proposes a Colony-RAN architecture for mobile networks. The novelty of Colony-RAN lies in the fact that by taking into account users' traffic demand, distribution and mobility, it can change the cell layout by dynamically adapting the connections between CUs and DUs. An optimization algorithm is presented by Carapellese et al¹³ for the CU placement problem over Fixed/Mobile Converged optical networks. The author formulates an ILP problem, which efficiently calculates the minimum number of CU pools, taking into account the maximum allowed distance between DUs and their CUs. The same author proposes an energy-efficient CU placement algorithm for optical networks¹¹, aiming to minimize the Aggregation Infrastructure Power. Using ILP techniques, Holm et al¹⁴ formulates a problem for optimizing cells assignments to different CU pools by taking into consideration fiber length and statistical multiplexing gain. Huang et al¹⁶ studies the CU placement problem considering the PHY-RF split and the intra-PHY-layer split at the base stations. An optimization problem is formulated having the objective of minimizing the total deployment cost of the mobile network while satisfying the traffic demands. Al-obaidi et al¹⁵ derives an analytical model aiming to find the optimal ratio between optical fibers and microwave links in the fronthaul network. Harutyunyan et al¹⁷ proposes a DU-CU mapping algorithm with the goal of minimizing the required investment in order to migrate from the traditional D-RAN to the C-RAN architecture. While Checko et al¹⁸ compares cost and energy consumption reduction in the C-RAN and in the traditional D-RAN, Ceragon Networks¹⁹ compares fiber-based and microwave-based fronthaul networks in terms of their deployment cost. It is concluded that in rural areas fiber-based fronthauls are more cost-effective for over distances less than 500m, while MMW-based fronthauls take the advantage at around 1.6 km and above. In urban areas, however, MMW links deployment is much more effective even for very short distances.

VNF Placement. Network Function Virtualization (NFV) is yet another technology that is expected to play a pivotal role in tackling the challenges of the present-day and the future 5G mobile networks²⁰. By decoupling network functions from their purpose-built proprietary hardwares and deploying Virtualized Network Functions (VNFs) on commodity servers, NFV is able to reduce both capital expenditure (CapEx) as well as operational expenditure (OpEx)²¹. Moreover, NFV is able to achieve more efficient utilization of mobile network resources by enabling coexistence of several logical VNF instances on the same hardware equipment. The fact that thanks to the NFV technology network functions can be represented as VNFs (i.e., softwares) poses the problem of deciding on where to place those VNFs such as to optimize the network resource utilization while satisfying the requirements of the VNFs. This problem is known as the VNF placement problem. The VNF placement and the datacenters/clouds component placement are conceptually similar. A number of studies have been published in this domain^{22,23,24,25}. A survey on resource management in cloud computing environments is presented by Jennings et al²⁶. Breitgand et al²² studies the problem of placing virtual machine instances on physical containers having an objective of reducing latency and communication overhead. A novel design is proposed²³ for components placement problem of scalable hierarchical applications in the cloud environment. In order to address the scalability problem of centralized algorithms, the authors propose a distributed algorithm that achieves performance similar to the one of centralized algorithms. The authors then extend their work²⁴ by proposing several algorithms aiming at efficiently managing data of component-based applications. Ghaznavi et al²⁵ jointly considers the trade-offs between bandwidth and host resource consumption, and the elasticity overhead in the VNF placement problem formulation. Guerzoni et al²⁷ and Despotovic et al²⁸ propose a joint node and link mapping algorithm, while Clayman et al²⁹, Moens et al³⁰ and Bari et al³¹ tackle the problem of dynamic VNF placement. Riggio et al³² studies the VNF placement problem for the radio access network. Mijumbi et al³³ formulates an online VNF scheduling and mapping problem proposing greedy algorithms and a tabu search-based heuristic, which are then compared in terms of their mapping cost, service processing time and revenue.

Functional Split. Several organization such as 3GPP⁴, NGMN³⁴, NGFI³⁵ are working on functional splits and the fronthaul technologies. RAN WG3 of 3GPP concentrates their work on the PDCP-RLC split (option 2) and the high RLC-low RLC

(option 3.1) split from the higher layer functional splits, and MAC–PHY split (option 6) and the intra–PHY (option 7) split from the lower–layer functional splits. According to them, the PDCP–RLC split is the most straightforward option to be standardized since the fundamentals for achieving it have already been standardized for LTE Dual Connectivity. NGMN considers only the splits within and up to the PHY layer, and their fronthaul requirements fully conform with the ones considered in our work. Whereas, NFGI proposes packet-based (Ethernet) fronthaul solutions, which is not the focus of our work. Table 2 shows different terminologies of RAN components and functional splits used by different organizations working on the functional splits and fronthaul technologies.

A detailed discussion on various functional splits has been conducted by Rost et al³⁶ and Wubben et al³⁷. Rost et al proposes the RAN as a Service (RANaaS) concept, which is characterized by centralized management (i.e., both full and partial centralization of RAN functionalities are possible) that can be adapted to the actual service demands. Whereas, Wubben et al introduces several functional splits and provides numerical results on the required backhaul data rates for each considered split option. Considering burstiness of mobile traffic and the fact that the traffic varies depending upon the area (e.g., residential, office) and time of a day, Checko et al³⁸ proposes mathematical and simulation methods for quantifying the multiplexing gain on a fronthaul network and a CU pool of the PHY–RF and the PDCP–RLC functional splits. Different from the aforementioned works, Arnold et al³⁹ discusses a control plane / user plane functional split, highlighting the pros and cons. Liu et al⁴⁰ presents a graph–based algorithm for analyzing different baseband functional splits, while Meader et al⁴¹ explores several wired/wireless fronthaul technologies as well as associated bandwidth and latency requirements for different functional splits.

The PHY–layer split allows inter–cell interference mitigation to be achieved through the implementation of advanced signal processing mechanisms, such as CoMP features (e.g., joint processing, coordinated scheduling/beamforming)⁷. A detailed investigation on the various functional splits at the PHY layer has been conducted by Dötsch et al⁴² and Bartelt et al⁴³.

Depending on the need, one functional split may be more beneficial over the other (see Section 2). Therefore, the possibility of flexibly selecting the functional split option can bring its benefits. Harutyunyan et al⁴⁴ presents an optimization problem to flexibly select the optimal functional split option for small–cell base stations having the goal of minimizing the network–wide inter–cell interference. Koutsopoulos et al⁴⁵ instead studies the problem of jointly optimizing the functional split selection and radio resource scheduling policy.

However, none of the mentioned works studies the dynamic CU placement problem over a reconfigurable substrate network with different functional split options. Moreover, none of them considers a reconfigurable MMW fronthaul, which based on certain criteria (e.g., daytime versus nighttime traffic patterns) can dynamically re–embed virtual networks with the objective of reducing the network–wide power consumption.

4 | NETWORK MODEL

The reference network architecture is depicted in Fig. 2. The lower part of the figure shows a traditional C–RAN deployment with the PHY–RF split where the CUs are centralized in powerful computing facilities, and CPRI optical fronthaul links are used to interconnect DUs with CU pools. The upper part of the figure instead shows the architecture envisioned in this work, where the CU pools and macro cells are co–located and connected to DUs through reconfigurable MMW fronthaul links. The reconfigurability of the MMW transmission links allows the interconnections between DUs to be reconfigured, aiming to minimize the number of active MMW connections. For example, when the traffic on a MMW link is low then, while embedding a new virtual network request, the algorithm may try to re–use the same MMW link, therefore, avoiding to power up a new MMW connection. Or, the algorithm may decide to shut down a low–utilized MMW link and aggregate its traffic in another more–utilized MMW link. A traditional S1 link is used for connecting CU pools with the core network. Compared to the traditional (expensive) CPRI optical links, the reconfigurable MMW fronthaul, although provides less fronthaul capacity, reduces the deployment and the power consumption costs of the fronthaul, while still allowing for improved control and coordination across the small cells.

In this section, we will first detail the notations used for the substrate and virtual network models. We will then introduce the optimal ILP formulation and a scalable heuristic algorithm for the dynamic CU placement problem. It is worthwhile to note that in this dynamic CU placement problem only MMW wireless fronthaul network is considered. Thus, the CU placement problem over the converged wireless/optical network is left as future work.

4.1 | Substrate Network Model

The considered substrate network is composed of computational as well as networking resources. The former consists of micro datacenters, possibly co-located with macro cells. The latter consists of a reconfigurable wireless fronthaul and multiple DUs. The fronthaul network consists of MMW routers equipped with a variable number of MMW interfaces and highly directional steerable antennas. Finally, each DU is equipped with a variable number of RF front-ends and with processing capabilities. We assume that DUs and MMW routers are co-located.

Let $G_s = (N_s, E_s)$ be an *undirected* graph modelling the substrate network, where $N_s = N_s^{du} \cup N_s^{cu}$ is the set of DUs sites/MMW Relays and CU pools, and E_s is the set of fronthaul links. Notice how N_s^{du} nodes⁴ in the substrate network can act as both DU and MMW Relays (i.e., they can both serve end-user terminals over, for example, an LTE air interface and act as relays in MMW fronthaul), while N_s^{cu} nodes (i.e., the ones co-located with macro cells) can act only as CU pools. A wireless edge $e^{nm} \in E_s$ if and only if a line-of-sight connection exists between the nodes $n, m \in N_s$.

Three weights, $\omega_a^s(n)$, $\omega_i^s(n)$ and $\omega_c^s(n)$, are assigned to each node $n \in N_s$: $\omega_{a,i,c}^s(n) \in \mathbb{N}^+$ representing, respectively, the number of RF front-ends, the set of MMW interfaces and the processing capacity available at the node n . Notice that, since different functional splits are considered for different embedding scenarios, the processing capacity of each substrate node (e.g., a DU, a CU pool) depends on the functional split option. We remind the reader that in the C-RAN architecture, a base station is decomposed into two parts, a DU and a CU, and the latter is centralized in a CU pool. The processing capacity at DUs and CU pools for the considered functional splits is reported in Table 6. Notice that the reported values are relative to the overall capacity of the substrate small cell n . For example, if the PHY-RF split is considered in the substrate network then the DUs do not possess processing capacity. Whereas, if the rest of the splits are considered, the DUs do possess processing capacity, which increases at the DUs and decreases at the CU pools when fewer upper-layer functionalities are centralized at the CU pools. Nevertheless, the overall processing capacity (the sum of the processing capacities at the CU pool and at the DUs) of the substrate network is equal to the number of substrate small cells and remain the same regardless of the considered functional split option (see Table 6). Each substrate node $n \in N_s$ is also associated with a geographic location $loc(n)$, as x, y coordinates, while each DU $n \in N_s^{du}$ is also associated with a coverage radius $\delta(n)$, in meters, indicating the coverage area of the small cell centered on DU n . Another weight $\omega_b^s(e^{nm})$ is assigned to each link $e^{nm} \in E_s$: $\omega_b^s(e^{nm}) \in \mathbb{N}^+$ representing the capacity (in Gbps) of the link connecting the two nodes $n, m \in N_s$. Finally, let P_s be the set of all loop-free substrate paths and $P_s(s, t)$ be the shortest path between $s, t \in N_s$. Table 3 summarizes the substrate network parameters.

A substrate network example is depicted in Fig.3a. This network composed of 21 nodes is representative of a scenario where DUs are deployed at road intersections in a dense Manhattan-like urban area. The considered topology is solely an example. The problem formulation is generic and can be used for other types of network topologies.

4.2 | Virtual Network Model

Virtual network requests are formulated as *undirected* graphs $G_v = (N_v, E_v)$, where $N_v = N_v^{du} \cup N_v^{cu}$ is the set of virtual DUs and virtual CU pools, and $E_v \subseteq N_v^{du} \times N_v^{cu}$ is the set of virtual fronthaul links. Notice that virtual network requests consist of virtual DUs and virtual CU pools, i.e., the requests do not contain MMW Relays. An edge $e^{nm} \in E_v$ if and only if the virtual DU n is connected to the virtual CU pool m . Thus, contrary to the substrate network model, edges in the virtual network requests represent the logical mapping between virtual DUs and their corresponding virtual CU pools. Additionally, we require that each virtual CU pool be mapped to one substrate CU pool. Conversely, different virtual CU pools can be mapped to the same substrate CU pool. This formulation allows MVNOs to specify requests in which a group of virtual CU pools is mapped to the same substrate CU pool, enabling advanced interference control algorithms such as Joint Transmission/Reception.

Nodes in the virtual network requests have two weights $\omega_c^v(n)$ and $\omega_a^v(n)$ denoting, respectively, the processing requirement and the number of RF front-ends requested by the node $n \in N_v$. Notice that the overall processing requirement of the small cell m in the virtual network request is the sum of the processing resource requirements of its components virtual DU $m_{du} \in N_v^{du}$ and virtual CU pool $m_{cu} \in N_v^{cu}$: $\omega_c^v(m) = \omega_c^v(m_{du}) + \omega_c^v(m_{cu})$. Each virtual DU $n \in N_v^{du}$ in the virtual network requests has a geographic location $loc(n)$ as x, y coordinates, which along with the location and the coverage radius of the substrate DUs is used to find the candidate substrate DUs for each requested virtual DU. Lastly, each virtual link $e^{nm} \in E_v$, which is the link connecting the two nodes $n, m \in N_v$ in virtual network requests, has a capacity requirement $\omega_b^v(e^{nm})$, which can be easily derived considering the characteristics of the cells and their employed functional split option. For example, given a 20 MHz FDD LTE

⁴Node is a general term used for both CU pools and DUs throughout the article.

cell with a 2x2 MIMO antenna configuration, and assuming that the cell employs the PHY–RF split, the required CPRI bitrate is ≈ 2.5 Gbps in order to achieve up to 150 Mbps of users traffic. Table 4 summarizes the virtual network request parameters.

Figure 3b displays an example of a virtual network request. The request is composed of three small cells. The number of requested small cells in the virtual network request is equal to the number of virtual DUs. Notice that the red ones require their CUs be co-located with the same CU pool, while the blue one does not impose such a constraint. Notice also that the blue small cell requests a single CU and a lower CPRI bandwidth compared to the red small cells.

4.3 | Fronthaul deployment cost analysis

In this section, we will perform a fronthaul deployment cost analysis studying the tipping point in terms of economic viability between an optical and a wireless fronthaul.

TCO of MNOs has been growing in the last few years in conjunction with exponentially increasing data traffic demands in mobile networks. The fronthaul network deployment cost constitutes the huge share of the TCO and, therefore, requires a careful design. Although, there are commercial products which provide more than 7.5 Gbps wireless CPRI links using E-band frequency range, the optical fiber, regardless of its cost, keeps leading the market as the best fronthaul medium in terms of its bandwidth and latency characteristics. However, apart from the cost factor, there are a number of other parameters, such as time-to-market, the physical feasibility of running wired/wireless links, which have to be taken into account when selecting the fronthaul medium. Therefore, future mobile networks are likely to leverage on both wired and wireless fronthauls.

Consider the network depicted in Fig.3a and assume that each small cell provides coverage within a radius of 250 m. Let us also assume that a MMW link can provide similar capacity as an optical fiber up to a distance of 500 m. The substrate network in this example has a total of 54 MMW interfaces for delivering connectivity from DUs to the CU pool; whereas, 18 km of optical fiber would be required to build the fronthaul network, like Checko et al^{46,18}, assuming that the optical fiber links are buried as straight lines from the CU pool to the DUs. Dividing the former value (54 MMW interfaces) by the latter one (18 km) we can obtain the cost ratio starting from which deploying the MMW links is more cost-effective than deploying fiber links. In this particular example, MMW links must be 3 times cheaper than one km of the optical fiber in order to make the MMW fronthaul economically viable. This cost factor reduces with an increase in the distance between the DUs and the CU pool.

5 | CU PLACEMENT

5.1 | Overview

Upon receiving a virtual network request, the InP that owns the substrate network has to decide whether to accept and map the request or to reject it. Efficient mapping of virtual network requests onto a substrate network is known as a virtual network embedding problem⁴⁷, which in the C–RAN is also called a CU placement problem. The problem is *NP*-hard and has been studied extensively in the literature^{28,48,49}. The embedding process is composed of two steps: the node embedding and the link embedding. In the node embedding step, each node in the virtual network request is mapped to a substrate node; while in the link embedding step, each link is mapped to a single substrate path. In both steps, nodes and links constraints must be satisfied.

One of the characteristics of MMW communication is that it requires highly directional antennas. This along with the fact that MMW Relays can possess only a few MMW interfaces makes only a subset of substrate links as viable to be used at a given time. As an example, let us consider the CU placement shown in Fig.3c. In this case, three small cells are embedded onto the substrate network. The red small cells are assumed to require two CPRI option 3 links (5 Gbps in total), which, assuming that each MMW interface has 5 Gbps capacity, results in a single MMW interface being required on the CU pool in order to serve the request, while two interfaces (i.e., one for serving the local small cell and one for relaying the CPRI link of the other small cells) being required on the relay DU. On the other hand, it is assumed that the blue small cell requires just a CPRI option 2 link (1.2 Gbps in total); therefore, a longer MMW link can be used, minimizing the number of relaying nodes required to serve the request. We will further discuss the relationship between CPRI capacity and MMW length in the evaluation section. Notice that in the alternative CU placement shown in Fig.3d, the constraint on the maximum number of interfaces of the relaying DU is violated since the mapping would require four interfaces while only two are actually available.

5.2 | ILP Formulation

Before formulating the ILP problem, we shall first find the candidate substrate DUs for each virtual DU in the virtual network request. This can be done by considering the location $loc(n')$ of the virtual DU $n' \in N_v^{du}$ along with the location $loc(n)$ and the coverage radius $\delta(n)$ of every substrate DU/Relay $n \in N_s^{du}$. For each virtual DU n' , a cluster of candidate DUs $\Omega(n')$ can then be defined as follows:

$$\Omega(n') = \left\{ n \in N_s^{du} \mid dis(loc(n), loc(n')) \leq \delta(n) \right\} \quad (1)$$

It is important to mention that the signal propagation model, although important, takes a secondary role in this CU placement problem. Firstly, this is because the CU placement problem only considers virtual network requests whose characteristics are specified by the MVNOs making the request. Thus, the actual users of MVNOs are not considered in the CU placement problem. Secondly, considering complex signal propagation models will unnecessarily complicate the embedding problem without adding any significant benefit. Therefore, the signal propagation of a cell is characterized by only its coverage radius.

We can now formulate the CU placement problem. This ILP formulation aims at computing the optimal CU placement by considering the available computational, fronthaul bandwidth and radio resources under a certain cost function with the goal of minimizing the power consumption in the fronthaul network. The chosen objective function is:

$$\text{minimize} \quad \sum_{e \in E_s} \sum_{e' \in E_v} \Lambda_e \omega_b^v(e') \Phi_e^{e'} + \sum_{n \in N_s} \sum_{e \in E_s(n)} \Lambda_i \Phi_e^i \quad (2)$$

where $\Phi_e^{e'} \in \{0, 1\}$ and $\Phi_e^i \in \{0, 1\}$ are binary variables that indicate, respectively, whether the virtual link $e' \in E_v$ has been mapped to the substrate link $e \in E_s$ and whether the MMW interface $i \in \omega_i^s(n)$ of the substrate link $e \in E_s(n)$ of the node $n \in N_s$ has been used in the virtual link mapping.

The first argument of the objective function minimizes the overall bandwidth consumption across all substrate fronthaul links. In other words, it minimizes the number of hops required to embed the virtual links onto the substrate paths. Whereas, the second argument minimizes the overall number of active MMW interfaces required to host the virtual network requests. It is important to mention that the cost for using the fronthaul bandwidth resources (Λ_e) is negligible compared to the cost of using MMW interfaces of the substrate nodes (Λ_i). This is because our goal is to reduce the power consumption in C-RAN by curtailing the number of active MMW interfaces.

The substrate network can embed virtual network requests as long as the host substrate nodes have enough processing resources, RF front-ends and fronthaul bandwidth in order to support the requests:

$$\sum_{n' \in N_v^{du}} \omega_c^v(n') \Phi_n^{n'} \leq \omega_c^s(n) \quad \forall n \in N_s^{du} \quad (3)$$

$$\sum_{n' \in N_v^{cu}} \omega_c^v(n') \Phi_n^{n'} \leq \omega_c^s(n) \quad \forall n \in N_s^{cu} \quad (4)$$

$$\sum_{e' \in E_v} \omega_b^v(e') \Phi_e^{e'} \leq \omega_b^s(e) \quad \forall e \in E_s \quad (5)$$

$$\sum_{n' \in N_v^{du}} \omega_a^v(n') \Phi_n^{n'} \leq \omega_a^s(n) \quad \forall n \in N_s^{du} \quad (6)$$

where the binary variable $\Phi_n^{n'} \in \{0, 1\}$ indicates whether the virtual node $n' \in N_v$ has been mapped to the substrate node $n \in N_s$. Specifically, constraints (3) and (4) ensure that the processing resource requested by the virtual DUs and the virtual CU pools are at most equal to the processing resource of, respectively, the host substrate DUs and CU pools. Constraint (5) makes sure that the overall fronthaul capacity of each substrate link is not exceeded. Finally, constraint (6) deals with the RF front-ends, making sure that the number of RF front-ends required at the host substrate node is at most equal to the maximum number of RF front-ends available at that substrate node.

Each node in the virtual network request must be mapped only once:

$$\sum_{n \in N_s} \Phi_n^{n'} = 1 \quad \forall n' \in N_v \quad (7)$$

Each virtual DU in the virtual network request must be mapped only on a substrate DU that belongs to its cluster of candidates:

$$\sum_{n \in N_s^{du} \setminus \Omega(n')} \Phi_n^{n'} = 0 \quad \forall n' \in N_v^{du} \quad (8)$$

At each substrate node, the sum of the substrate MMW interfaces that are used to carry the traffic of the virtual links must be less or equal to the number of MMW interfaces available at that node:

$$\sum_{e^{ij} \in E_v} \Phi_{e^{nm}}^{e^{ij}} + \sum_{e^{ij} \in E_v} \Phi_{e^{nm}}^{e^{ij}} \leq |\omega_i^s(n)| \quad \forall n, m \in N_s, \quad \forall e^{nm} \in E_s \quad (9)$$

The following constraint enforces for each virtual link $e^{nm} \in E_v$ to be a continuous path between the pair of physical nodes on top of which the virtual nodes $n, m \in N_v$ have been mapped:

$$\sum_{e \in E_s^{*i}} \Phi_e^{e^{nm}} - \sum_{e \in E_s^{i*}} \Phi_e^{e^{nm}} = \begin{cases} -1 & \text{if } i = n \\ 1 & \text{if } i = m \\ 0 & \text{otherwise} \end{cases} \quad \forall i \in N_s, \quad \forall e^{nm} \in E_v \quad (10)$$

where E_s^{*i} is the set of the fronthaul links that originate from any node and directly arrive at the node $i \in N_s$, while E_s^{i*} is set of the fronthaul link that originate from the node $i \in N_s$ and arrive at any node directly connected to i .

Finally, in order to minimize the number of MMW interfaces (i.e., second argument in the objective function) that are being used to deliver connectivity from the host DUs to the host CU pools, the last constraint guarantees that the MMW interfaces can be reused. Notice that the interface $i \in \omega_i^s(n)$ of the node $n \in N_s$ is used in mappings (i.e., $\Phi_e^i = 1$) if at least one virtual link has been mapped onto the substrate path that is using the interface i .

$$\sum_{e' \in E_v} \Phi_e^{e'} - \mu_b \Phi_e^i \leq 0 \quad \forall n \in N_s, \quad \forall i \in \omega_i^s(n), \quad e = E_s(n) \quad (11)$$

where μ_b is a big positive number. If the interface i has not been used to map a virtual link ($\sum_{e' \in E_v} \Phi_e^{e'} = 0$) then that interface will not be selected ($\Phi_e^i = 1$ is excluded) since the objective function, apart from minimizing the fronthaul bandwidth consumption, also aims at minimizing the number of active MMW interfaces.

The ILP-based placement algorithm has a limited scalability and, therefore, it is not applicable to big-sized networks. For example, it can take around one day to embed a virtual network request that has 4 nodes (1 CU pool and 3 DUs) over a grid-shaped substrate network composed of 49 nodes (2 CU pools and 47 DUs) on Intel Core i7 laptop (3.0 GHz CPU, 16 Gb RAM) using the Matlab[®] ILP solver (intlinprog). This is because the ILP-based algorithm considers all possible embedding solutions in order to find the optimal one. In this section, a heuristic algorithm is presented that is able to embed similar requests and find near-optimal embedding solutions in less than 10 milliseconds.

5.3 | Heuristic

The proposed heuristic algorithm consists of three steps (see pseudocode in Alg. 1). Let $m_{du} = |N_s^{du}|$ and $m_{cu} = |N_s^{cu}|$ be the number of, respectively, substrate DUs and substrate CU pools with $m = m_{du} + m_{cu}$. Similarly, let $n_{du} = |N_v^{du}|$ and $n_{cu} = |N_v^{cu}|$ be the number of, respectively, virtual DUs and virtual CU pools. Lastly, let $k = |E_s|$ be the number of edges available in the substrate network.

In the first step, the heuristic selects a list of candidate substrate DUs and candidate substrate CU pools, respectively, for each virtual DU $n \in N_v^{du}$ and each virtual CU pool $m \in N_v^{cu}$ in the virtual network request. Specifically, the candidate substrate DUs are selected based on the locations of the substrate and the virtual DUs, and based on the availability of the required RF front-ends and the processing resources at the substrate DUs (lines from 3 to 12 in the pseudocode). Whereas, the candidate substrate CU pools are selected by considering only the processing resource requirement of the virtual CU pools (lines from 13 to 19 in the pseudocode). Step 1 requires $O(n_{du}m_{du} + n_{cu}m_{cu})$ time.

In the second step, for each candidate substrate CU pool $p \in candidates(n)$ of each virtual CU pool $n \in N_v^{cu}$, the heuristic considers all neighbor nodes (i.e., the rest of the nodes in the virtual network request) of the virtual node n . Then, the heuristic computes the cost of embedding each virtual node pair n, m . In essence, this is the cost of embedding the virtual link e^{nm} onto the path between the candidate CU pool $p \in candidates(n)$ and the candidate DU $q \in candidates(m)$. The heuristic then assigns the virtual node n to the substrate node p with the lowest mapping cost (lines from 21 to 37 in the pseudocode). Here, the goal is to place the virtual CU pool on the substrate CU pool that can support all its virtual DUs at the minimal cost. This process requires $O(n_{cu}m_{cu}n_{du}(m_{du} - 1)k \log_{10} m)$ time.

In the last step, the heuristic loops over the virtual CU pools, considers the neighbor nodes $neighbor(n)$ (i.e., the virtual DUs) of each virtual CU pool $n \in N_v^{cu}$ then maps each virtual DU $m \in neighbor(n)$ to its candidate substrate DU $q \in candidates(m)$ that has the lowest mapping cost. Once the virtual DU has been mapped, using Dijkstra's shortest path algorithms,

Algorithm 1 Nodes and links assignment

```

1: procedure EmbedRequest( $G_s, G_v$ )
2:   Step 1: Compute list of candidates.
3:   for  $n \in N_v^{du}$  do ▷ Virtual DUs.
4:     for  $p \in N_s^{du}$  do ▷ Substrate DUs.
5:        $d \leftarrow \text{dis}(\text{loc}(n), \text{loc}(p))$  ▷ Distance in meters.
6:       if  $d \leq \delta(p)$  then
7:         if  $\omega_a^v(n) \leq \omega_a^s(p)$  and  $\omega_c^v(n) \leq \omega_c^s(p)$  then
8:            $\text{candidates}(n) \leftarrow p$ 
9:         end if
10:      end if
11:    end for
12:  end for
13:  for  $m \in N_v^{cu}$  do ▷ Virtual CU pools.
14:    for  $q \in N_s^{cu}$  do ▷ Substrate CU pools.
15:      if  $\omega_c^v(m) \leq \omega_c^s(q)$  then
16:         $\text{candidates}(m) \leftarrow q$ 
17:      end if
18:    end for
19:  end for
20:  Step 2: Perform CU placement.
21:  for  $n \in N_v^{cu}$  do ▷ Virtual CU pools.
22:    for  $i \in N_s$  do ▷ Initialize mapping cost array.
23:       $m_c(i) \leftarrow 0$ 
24:    end for
25:    for  $p \in \text{candidates}(n)$  do
26:      for  $m \in \text{neighbors}(n)$  do
27:         $c_{curr} \leftarrow +\infty$ 
28:        for  $q \in \text{candidates}(m)$  do
29:           $c_{new} \leftarrow \sum_{e \in P_s(p,q)} \omega_e^v(e^{nm})$ 
30:           $c_{curr} \leftarrow \min(c_{curr}, c_{new})$ 
31:        end for
32:         $m_c(p) \leftarrow c_{curr}$  ▷ Accumulate mapping cost.
33:      end for
34:    end for
35:     $p \leftarrow \text{argmin}(m_c(p))$ 
36:     $\text{mapped}(n) \leftarrow p$ 
37:  end for
38:  Step 3: Perform RF front-ends embedding.
39:  for  $n \in N_v^{cu}$  do ▷ Virtual CU pools.
40:     $p \leftarrow \text{mapped}(n)$ 
41:    for  $m \in \text{neighbors}(n)$  do
42:      for  $i \in N_s$  do ▷ Initialize mapping cost array.
43:         $m_c(i) \leftarrow 0$ 
44:      end for
45:      for  $q \in \text{candidates}(m)$  do
46:         $m_c(q) \leftarrow \sum_{e \in P_s(p,q)} \omega_e^v(e^{nm})$ 
47:      end for
48:       $q \leftarrow \text{argmin}(m_c(q))$ 
49:       $\text{mapped}(m) \leftarrow q$ 
50:      Allocate path  $P_s(p, q)$ 
51:    end for
52:  end for
53: end procedure

```

the heuristic allocates the shortest path $P_s(p, q)$ between the nodes $p, q \in N_s$, which are ones that have hosted, respectively, the virtual CU pool n and the virtual DU m (lines from 39 to 50 in the pseudocode). This results in the virtual nodes of the virtual network request being placed close to each other in the substrate network and, therefore, entails to an efficient utilization of the substrate resources. Step 3 requires $O(n_{cu}n_{du}(m_{du} - 1)k \log_{10} m)$ time. Thus, the overall time complexity of the heuristic is $O(n_{du}m_{du} + n_{cu}m_{cu} + [n_{du}n_{cu}(m_{du} - 1)k \log_{10} m](1 + m_{cu}))$.

Finally, it is important to mention that, in order to ensure their correctness of the results, we pass all the solutions found by the heuristic through the same constraints defined for the ILP formulation.

6 | EVALUATION

In this section, we compare⁵ the performance of the ILP-based placement algorithm with the performance of the heuristic using different synthetic substrate networks and different virtual network requests. Initially, we describe the used simulation environment and the performance metrics. We then report on the outcomes of the numerical simulations conducted in a discrete event simulator implemented in Matlab.

6.1 | Simulation Environment

The simulation parameters and the substrate network characteristics are derived from a number of works on MMW communications. Rappaport et al⁵⁰ suggests that optimum coverage be achieved by having 200 m of distance between each DU. Pi et al⁵¹ estimates 1 km to be the typical coverage radius for MMW links in line-of-sight conditions. Finally, Ghosh et al⁵² and Rangan et al⁵³ rely on empirical measurements to show that bitrates as high as 10 Gbps can be achieved in the MMW band with an outage probability of $\approx 11\%$, while 5 Gbps of bitrate can be achieved with an outage probability of $\approx 3\%$.

The ILP-based algorithm and the heuristic are evaluated using grid-shaped and random substrate network topologies. The former is similar to the one depicted in Fig.3a and is composed of 25 nodes with a uniform inter-node distance of 500 m. Whereas the latter, as the name implies, is generated by randomly positioning the same number of nodes in the area of 4 km². As opposed to the grid-shaped substrate network topology, in the random substrate network topology the nodes may have more/fewer neighbors and, therefore, may also have more/fewer link embedding opportunities. The random network is generated with the goal of mimicking the real-life mobile network deployment scenarios such as mobile networks deployed in office or residential areas. In order to make a fair comparison between those two topologies, the total number of MMW interfaces constant across the two types of substrate network topologies is kept the same. In both cases, the following functional splits are considered: PHY-RF split, PHY split, MAC split, and PDCP-RLC split⁶.

When generating the substrate topologies, it is assumed that there is a line-of-sight between substrate nodes and that up to 500 m of inter-node distance the link can support up to 2.5 Gbps of traffic, while up to 1000 m of inter-node distance the link can deliver up to 1.2 Gbps of traffic. Table 5 summarizes some of the most common CPRI options for the PHY-RF split. It can be observed that the aforementioned assumptions correspond to a CPRI option 3 and to a CPRI option 2 configurations, respectively.

Each node in the substrate networks is either a DU/Relay or a CU pool. However, depending upon the functional split used in the substrate network, both of them can possess processing units. In the case of the PHY-RF split in the substrate network, for example, the DU $n_{du} \in N_s^{du}$ of the substrate small cell n does not possess processing capacity, while in the case of the PHY split or the MAC split, the DU does possess processing capacity $\omega_c^s(n_{du}) = 0.5 \cdot \omega_c^s(n)$ and $\omega_c^s(n_{du}) = 0.7 \cdot \omega_c^s(n)$, respectively, where $\omega_c^s(n)$ is the overall processing capacity of the substrate small cell n . Notice that, for example, in the case of the PHY split, it is assumed that the half of the processing capacity is allocated to the DUs while the other half is allocated to the CU pools. This is because the most processor-hungry procedure (i.e., FFT/IFFT) is taking place in the PHY layer. The processing requirement increases at the DUs and decreases at the CU pools when fewer layers (e.g., PHY layer, MAC layer) are centralized at the CU pools (see Table 6). The overall processing capacity each substrate network is computed considering the number of embeddings, the average number of DUs in a virtual network request and the functional split option employed in the substrate network.

⁵Note that we do not compare the proposed algorithm with any of the state-of-the-art algorithms since the problem formulations are different and the comparison would be unfair.

⁶For the sake of improving readability, only the results of the first three splits are reported. The results of the PDCP-RLC split resemble the ones of MAC split and are thus omitted.

The number of CU pools, which are randomly deployed over the substrate networks, varies between 1 and 4. However, for the sake of fair comparison, the same quantity of CU pools with the same locations are selected for virtual network embedding problems for different substrate networks with different functional splits. DU relays at the edges of the network are equipped with a single MMW interface and 4 RF front-ends. Whereas, the rest of the DUs are equipped with 4 MMW interfaces and 8 RF front-ends. Lastly, CU pools are equipped with 10 MMW interfaces and do not possess RF front-ends.

Virtual network requests consist of star-shaped networks similar to the ones shown in Fig.3b. The number of small cells per virtual network request and the number of RF front-ends per small cell are randomly selected within the set of $\{1, 2, 3, 4\}$ and $\{1, 2\}$, respectively. Note that these small sets are selected only for the sake of comparing the ILP-based algorithm with its heuristic counterpart, which is far more scalable as discussed in Section 6.2. The fronthaul bandwidth requirement instead depends on the functional split of the virtual network request. For example, if the PHY-RF split is considered, each of the RF front-ends may require either a CPRI option 2 or a CPRI option 3 link. The overall processing requirement of a virtual network request is equal to the number of requested DUs. Whereas, the processing requirements of individual nodes (i.e., a virtual DU or a virtual CU pool) depend upon the considered functional split. For example, if the PHY split or the MAC split are considered, the processing requirements of the virtual DU $n_{du} \in N_v^{du}$ would be, respectively, $\omega_c^v(n_{du}) = 0.5 \cdot \omega_c^v(n)$ and $\omega_c^v(n_{du}) = 0.7 \cdot \omega_c^v(n)$ where $\omega_c^v(n)$ is the overall processing requirement of the virtual small cell n . Whereas, the remaining $\omega_c^v(n_{cu}) = 0.5 \cdot \omega_c^v(n)$ and $\omega_c^v(n_{cu}) = 0.3 \cdot \omega_c^v(n)$ would be the processing requirements of the virtual CU pool $n_{cu} \in N_v^{cu}$, respectively, in the PHY split and the MAC split cases. The fronthaul bandwidth requirements for different antenna configurations as well as the processing resource requirements at both DUs and CU pools for the considered functional splits can be found in Table 6. Note that the fronthaul bandwidth requirements are derived for a 20 MHz LTE channel. More details can be found in SCF⁹.

In this study, it is assumed that a fixed number of virtual network requests are embedded sequentially. It is important to mention that with the arrival of a new virtual network request, all the previously embedded requests are re-computed along with the new request (global network optimization). The re-embedding can also be triggered based on the daytime versus the nighttime traffic requirement variation. Thus, a dynamic virtual network embedding is considered.

In order to make a fair comparison between the different splits, the same virtual network requests are used for all the functional split cases. Specifically, 200 (10 simulations each with 20 embeddings) random virtual network requests are generated and used by both the ILP-based algorithm and the heuristic for different functional split cases in the grid-shaped and the random substrate networks. Reported results are the average of 10 simulations.

6.2 | Simulation Results

Figure 4 displays the acceptance ratio, the embedding cost and the execution time of the ILP-based CU placement algorithm and of the placement heuristic for different functional splits in the grid-shaped⁷ substrate network. The acceptance ratio is computed in two ways: based on the number of accepted virtual network requests (see Fig.4a) and based on the number of hosted RF front-ends (see Fig.4b). We remind the reader that virtual network requests can have a variable number of DUs each requesting a variable number of RF front-ends. As expected, in both figures the ILP-based algorithm achieves a higher acceptance ratio compared to the one of the heuristic in all functional split cases.

In Fig.4a, a saturation point in the acceptance ratio can be observed for the PHY-RF split when the number of CU pools in the substrate network is three. Initially, when there is one CU pool in the substrate network, the acceptance ratio for both algorithms is low (65%) for the PHY-RF split since compared to the rest of the splits it has much higher fronthaul bandwidth requirement (see Table 1), which becomes a bottleneck for accepting more virtual network requests. Since, as it is mentioned in Subsection 6.1, CU pools have more MMW interfaces than DUs, deploying more CU pools in the substrate network increases the network-wide fronthaul bandwidth capacity, which leads to a higher acceptance ratio. We can observe, however, that the acceptance ratio for the PHY-RF split reduces when the number of CU pools is four. This can be explained by the fact that at this point the RF front-ends start to become a bottleneck for accepting more request. Because of the same reason, we can also observe that the acceptance ratio for the rest of the splits reduces when there is more than one CU pool in the substrate network. Notice that as opposed to the PHY-RF split, the fronthaul capacity never becomes a bottleneck for the PHY and MAC splits in order to accept more virtual network requests. This is justified by the fact that compared to the PHY-RF split those splits have significantly lower fronthaul bandwidth requirements (see Table 1).

⁷The performance results of the random topology are similar to the ones of the grid topology.

It is interesting to note that, even though the acceptance ratio for the PHY–RF split (see Fig.4a) decreases when four CU pools are available in the substrate network, the acceptance ratio in terms of the on–boarded RF front–ends (see Fig.4b) increases for the same split for the same number of CU pools in the substrate network. This is because substituting more DU nodes with CU pools curtails the overall number of DU nodes in the substrate network, which in turn curtails the number of RF front–ends that can host requested RF front–ends, since in the considered scenario only DUs possess RF front–ends. As a result, the remaining substrate DUs become heavily utilized and, therefore, the acceptance ratio of the RF front–ends increases across the entire substrate network. Notice also that there is a significant difference between the acceptance ratios in Fig.4a and Fig.4b. This is because the substrate DU nodes possess redundant RF front–ends in order to support the extreme case in which each virtual DU of the virtual networks might request two RF front–ends, which is the maximum number of RF front–ends that a virtual DU can request. As mentioned, however, each virtual DU selects the number of RF front–ends randomly between one and two.

Figure 4c plots the average embedding cost that is defined as the average number of MMW interfaces that are used to map a virtual link onto a substrate path. It can be observed that the ILP–based embedding algorithm can on–board more requests than the heuristic, and at the same time, ensure a lower embedding cost. This means that the ILP–based algorithm employs the MMW interfaces more efficiently than the heuristic. It can also be observed that for all the functional splits, the embedding cost demonstrates an increasing trend with more number of CU pools in the substrate network. This is due to the fact that compared to the DUs, CU pools possess more MMW interfaces, which leads to more link embedding opportunities, therefore increasing the embedding cost. Lastly, it can be seen that the embedding cost for the PHY–RF split is always higher compared to the cost of the rest of the splits, regardless of the quantity of the CU pools in the substrate network. Similarly, the embedding cost for the PHY split is always higher than the one of the MAC split. This can be explained by the fact that the higher–layer is the functional split, the less is the fronthaul bandwidth requirement, which results in more virtual links being mapped to the same MMW interface. Therefore, the lower–layer is the functional split, the cheaper is the MMW links to be used.

Figure 4d illustrates the average amount of time required in order to embed a *single* virtual network request by using the ILP–based placement algorithm and the placement heuristic. It can be seen that the embedding time for the ILP–based algorithm is significantly higher compared to the embedding time of the heuristic. The problem with the ILP–based algorithm is that it becomes computationally intractable when a substrate network with a few tens of nodes is considered, while the heuristic is able to effectively map complex virtual network requests on the substrate network with a few hundreds of nodes in a limited amount of time. Even though for the sake of finding the optimal CU placement MNOs may willingly wait even several weeks, we argue that the proposed heuristic, although less efficient than the ILP–based placement algorithm, allows a faster service on–boarding time, while the network configuration could be periodically optimized by using the ILP–based placement algorithm.

Figure 5 shows the resource utilizations of the substrate network for the ILP–based algorithm and the heuristic. Specifically, Fig.5a and Fig.5b illustrate the processing resource utilization of substrate DUs and CU pools, respectively. In both figures, it can be observed that the resource utilization of the ILP–based algorithm is higher than the one of the heuristic. While the processing resource utilization at the CU pools (see Fig.5b) resembles the acceptance ratio (see Fig.4a), the picture is the totally different in the processing resource utilization at the DUs (see Fig.5a). For the PHY–RF split, the DU utilization is zero since in that case, the DUs do not possess processing resource. For the other two splits instead, the DU utilization as well as RF front–end utilization (see Fig.5c) increases with an increase in the number of CU pools in the substrate network. This is justified by the fact that increasing the number of CU pools in the substrate network curtails the overall number of DU nodes in the network, therefore, the remaining DUs become much more loaded.

Figure 5d shows the network–wide MMW⁸ interface utilization. It can be observed that the utilization increases for all the splits for both algorithms with the increase in the number of CU pools. It can also be observed that compared to the heuristic algorithm, in the case of using the ILP–based algorithm, the MMW interface utilization for the PHY and MAC splits is lower, regardless of the number of CU pools. This is an evidence of the fact that the ILP–based algorithm exploits the MMW interfaces in a more efficient manner than the heuristic. The picture is similar for the PHY–RF split only when there is one CU pool in the substrate network. When there are more than one CU pool in the substrate network, the MMW interface utilization is more in the case of the ILP–based algorithm compared to its heuristic counterpart since, as it can be seen in Fig.4a, the ILP–based algorithm accepts a significantly more number of virtual network requests. It is worthwhile to note that although the acceptance ratio increases when less resources are centralized (i.e., from the PHY–RF split towards the MAC split), the MMW interface utilization reduces due to the fact that the MMW interfaces that are already in use are being reused more frequently, avoiding to power up new MMW interfaces.

⁸The evaluation of the MMW connection is out of the scope of this work.

In order to compare different substrate topologies, let us analyze Fig.6a and Fig.6b that show the RF front–end and interfaces utilization ratios for, respectively, the grid–shaped and the random substrate networks. The higher is the ratio, the more beneficial the embedding is. In other words, more interfaces are being reused to map virtual network requests. The superiority of the ILP–based algorithm over the heuristic can be observed in terms of a higher ratio in both grid and random substrate topologies. The results in both plots are somewhat similar in terms of their values and fluctuating behavior. On the one hand, increasing the number of CU pools in the substrate network increases the utilization of the RF front–ends because of the lack of DUs. On the other hand, it increases the interface utilization up to its saturation point (i.e., when the number of CU pools is three). Therefore, the optimal RF front–ends/MMW interfaces ratio varies and depends on the actual functional split used at the RAN.

In order to better understand how the substrate resources are utilized during the embedding process, let us now analyze a single iteration of the simulation in detail. For the sake of improving the readability, only two splits are shown: the PHY split and the MAC split for the DU utilization figures, while the PHY–RF split and the MAC split for the rest of the figures. We remind the reader that each iteration consists of 20 randomly generated virtual network requests. As it can be seen in Fig. 7, the utilization of the substrate resources in the case of employing the ILP–based placement algorithm is higher compared to the resource utilization of the heuristic, regardless of the number of CU pools in the substrate network. This is justified by the fact that the ILP–based placement algorithm is able to embed more requests than the heuristic.

Figures 7a, 7b, 7c and 7d show the acceptance ratio for a different number of CU pools. This ratio is calculated as the percentage of the accepted requests out of 20 virtual network requests (a single iteration). As expected, the acceptance ratio of the PHY–RF split is smaller than the acceptance ratio of the MAC split for both algorithms, regardless of the number of CU pools in the substrate network. This is because the latter split has 16.5 times smaller fronthaul bandwidth requirement compared to the former one (see Table 1). Thus, the higher–layer splits, such as the MAC split, are always beneficial in terms of fronthaul resource requirements and, therefore, are more energy–efficient and enable more virtual network requests to be embedded in the substrate network. We can observe that for all the splits the ILP–based algorithm has accepted more number of virtual network requests than its heuristic counterpart.

Figures 7e, 7f, 7g and 7h resemble Fig.4b and Fig.5a. The DU utilization increases with an increase in the number of CU pools. As we have already mentioned, the reason for this is that the number of CU pools is increased at the expense of the number of DUs in the substrate network. Figures 7i, 7j, 7k and 7l show the utilization of the CU pools. It is worthwhile to notice that, for example, when there is one CU pool in the substrate network, both the ILP–based placement algorithm and the placement heuristic embed all the virtual network requests up to 10 for all the splits. Therefore, the DU and the CU pool utilizations are the same for all of them, although, the CU pool processing requirements of the PHY–RF split and the MAC split are different. This is because the same requests, with different processing requirements, are used for different functional split cases, and the processing resources of the CU pools are picked based on the split option that the embedding is considered for. We can observe that the CU pool utilization pattern resembles the acceptance ratio of the virtual network requests (see Fig.4a). Initially, when there is only one CU pool in the substrate network, the CU pool utilization for the MAC split for both algorithms is much higher compared to the ones of the PHY–RF split. This can be explained by the fact that compared to the PHY–RF split, when the MAC split is employed, many more virtual network requests are accepted by the substrate network. As the number of CU pools increases up to three, the gap in the CU pool utilization between the MAC split and the PHY–RF split shrinks for both algorithms. This is due to the fact that the acceptance ratio increases for the PHY–RF split and decreases for the MAC split when two or three CU pools are deployed in the substrate network. Whereas, when there are four CU pools in the substrate network, this gap increases since apart from the MAC split, the acceptance ratio also reduces for the PHY–RF split as a result of insufficient substrate RF front–ends. A similar behavior can be observed in the RF front–ends utilization graphs (see Figs. 7m, 7n, 7o and 7p).

Finally, Figs. 7q, 7r, 7s and 7t plot the utilization of MMW interfaces for one iteration. When the number of CU pools in the substrate network is one or two, the difference in the acceptance ratio for both splits for the ILP–based algorithm is significantly higher than for the heuristic. Therefore, in those cases, the ILP–based algorithm has a higher MMW interface utilization ratio than the heuristic. As opposed to the previous case, when the number of CU pools in the substrate network is three or four, the difference in the acceptance ratio of both splits between the ILP–based algorithm and the heuristic is negligible. Therefore, in those cases, the ILP–based algorithm has a lower MMW interface utilization ratio than the one of the heuristic. This means that the ILP–based algorithm is able to utilize the MMW interfaces more efficiently than its heuristic counterpart.

7 | CONCLUSIONS

Small cells are becoming more and more ubiquitous in mobile networks as a cheap solution to boost the system capacity of the currently available and forthcoming mobile networks. However, in order to make their widespread adoption economically more viable, flexible fronthaul technologies such as wireless fronthauling are required.

In this paper, we provided a formulation for the dynamic CU placement problem for mobile networks with different functional splits. We introduced an ILP-based placement algorithm in order to solve the placement problem for small-scale networks and a CU placement heuristic for large-scale networks. After performing extensive simulations, we can conclude that an optimal functional split selection can lead to significant benefits in terms of MMW interface utilization, fronthaul bandwidth utilization and processing resource utilization at both DUs and CU pools since there is a large difference in the aforementioned parameters for the considered splits. We can also conclude that the higher-layer is the functional split option, the more efficient is the utilization of the substrate resources, although, the less is the resource centralization benefits. Thus, MNOs should tend to use higher-layer functional splits at their C-RAN. Although, the optimal functional split selection for the MNOs depends on a number of parameters, and sometimes there might be a need for using, for example, the PHY-RF split in order to be able to exploit advanced interference cancellation/avoidance techniques, at the expense of inefficient utilization of the substrate resources.

With regard to CU pools in the substrate network, two of the important factors that in our opinion should be taken into account when deploying CU pools are the distance between CU pools and DUs, and the deployment cost of CU pools such as to minimize the overall C-RAN deployment cost, avoid underutilization of CU pools and meet the traffic demand at DUs connected to CU pools at any time.

As future work, our plan is to consider more complex scenarios by extending the problem formulation. In particular, we want to consider a real mobile network in which both wireless and optical links are used as a transport medium. Based on the available transport medium option as well as real traffic requirements, we want to study an optimal functional split selection for each base station.

References

1. *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016-2021*. : Cisco; 2017.
2. *Ericsson Mobility Report, 2016-2021*. : Ericsson; June, 2017.
3. Chen K., Duan R.. C-RAN The Road Towards Green RAN, Version 2.5. White Paper, 2011;.
4. *Technical Specification Group Radio Access Network; Study on New Radio Access Technology; Radio Access Architecture and Interfaces*. rep. tr 38.801 v2.0.0: 3GPP Sophia Antipolis, France; 2017.
5. Larsen Line MP, Checko Aleksandra, Christiansen Henrik L. A survey of the functional splits proposed for 5G mobile crosshaul networks. *IEEE Communications Surveys & Tutorials*. 2018;.
6. *Common Public Radio Interface, Interface Specification, V7.0*. : CPRI Bengaluru, India; 2015.
7. Sesia Stefania, Toufik Issam, Baker Matthew. *LTE - the UMTS long term evolution: from theory to practice*. Wiley; 2011.
8. Riggio Roberto, Harutyunyan Davit, Bradai Abbas, Kuklinski Slawomir, Ahmed Toufik. SWAN: Base-Band Units Placement over Reconfigurable Wireless Front-Hauls. In: Proc. of IEEE CNSM; 2016; Montreal, Quebec, Canada.
9. *Small cell virtualization functional splits and use cases, V7.0*. rep. 159.07.02: Small Cell Forum London, U.K; 2016.
10. Checko Aleksandra, Christiansen Henrik L, Yan Ying, et al. Cloud ran for mobile networks—a technology overview. *Communications Surveys & Tutorials, IEEE*. 2014;17(1):405–426.
11. Richter Fred, Fehske Albrecht J, Fettweis Gerhard P. Energy efficiency aspects of base station deployment strategies for cellular networks. In: Proc. of IEEE VTC 2009-Fall; 2009; Anchorage, AK, USA.
12. Namba Shinobu, Matsunaka Takashi, Warabino Takayuki, Kaneko Shoji, Kishi Yoji. Colony-RAN architecture for future cellular network. In: Proc. of IEEE FutureNetw; 2012; Berlin, Germany.

13. Carapellese Nicola, Tornatore Massimo, Pattavina Achille. Placement of base-band units (BBUs) over fixed/mobile converged multi-stage WDM-PONs. In: Proc. of IEEE ONDM; 2013; Brest, France.
14. Holm Henrik, Checko Aleksandra, Al-obaidi Rami, Christiansen Henrik. Optimal assignment of cells in C-RAN deployments with multiple BBU pools. In: Proc. of EuCNC; 2015; Paris, France.
15. Al-obaidi Rami, Checko Aleksandra, Holm Henrik, Christiansen Henrik. Optimizing Cloud-RAN deployments in real-life scenarios using Microwave Radio. In: Proc. of EuCNC; 2015; Paris, France.
16. Huang Bo-Syuan, Chiang Yi-Han, Liao Wanjiun. Remote radio head (RRH) deployment in flexible C-RAN under limited fronthaul capacity. In: Proc. of IEEE ICC ; 2017; Paris, France.
17. Harutyunyan Davit, Riggio Roberto. How to Migrate From Operational LTE/LTE-A Networks to C-RAN With Minimal Investment?. *IEEE Transactions on Network and Service Management*. 2018;15(4):1503–1515.
18. Aleksandra Checko, Christiansen Henrik Lehrmann, Berger Michael StÅijbert. Evaluation of energy and cost savings in mobile Cloud RAN. In: Proc. of OPNETWORK; 2013.
19. *Mobile backhaul: Fiber vs. microwave.* : Ceragon Networks; 2009.
20. Nguyen Van-Giang, Do Truong-Xuan, Kim YoungHan. SDN and virtualization-based LTE mobile network architectures: A comprehensive survey. *Wireless Personal Communications*. 2016;86(3):1401–1438.
21. ETSI Network Function Virtualization (NFV): Architectural Framework. ETSI GS NFV 002 v1.1.1 (2013-10) [Online]. Available: https://www.etsi.org/deliver/etsi_gs/nfv/001_099/002/01.01.01_60/gs_nfv002v010101p.pdf.
22. Breitgand D., Epstein A., Glikson A., Israel A., Raz D.. Network aware virtual machine and image placement in a cloud. In: Proc. of IEEE CNSM; 2013; Zurich, Switzerland.
23. Barshan M., Moens H., De Turck F.. Design and evaluation of a scalable hierarchical application component placement algorithm for cloud resource allocation. In: Proc. of IEEE CNSM; 2014; Rio, Brasil.
24. Barshan M., Moens H., Latre S., De Turck F.. Algorithms for efficient data management of component-based applications in cloud environments. In: Proc. of IEEE NOMS; 2014; Krakow, Poland.
25. Ghaznavi M., Khan A., Shahriar N., Alsubhi K., Ahmed R., Boutaba R.. Elastic Virtual Network Function Placement. In: Proc. of IEEE CloudNet; 2014; Niagara Falls, Canada.
26. Jennings Brendan, Stadler Rolf. Resource management in clouds: Survey and research challenges. *Journal of Network and Systems Management*. 2015;23(3):567–619.
27. Guerzoni R., Trivisonno R., Vaishnavi I., et al. A novel approach to virtual networks embedding for SDN management and orchestration. In: in Proc. of IEEE NOMS; 2014; Krakow, Poland.
28. Despotovic Zoran, Hecker Andreas, Malik Aqdas Naveed, et al. VNetMapper: A fast and scalable approach to virtual networks embedding. In: Proc. of IEEE ICCCN; 2014; Shanghai, China.
29. Clayman S., Maini E., Galis A., Manzalini A., Mazzocca N.. The dynamic placement of virtual network functions. In: Proc. of IEEE NOMS; 2014; Krakow, Poland.
30. Moens H., De Turck F.. VNF-P: A model for efficient placement of virtualized network functions. In: Proc. of IEEE CNSM; 2014; Rio, Brasil.
31. Bari Faizul, Chowdhury Shihabur Rahman, Ahmed Reaz, Boutaba Raouf. On Orchestrating Virtual Network Functions. In: Proc. of IEEE CNSM; 2015; Barcelona, Spain.
32. Riggio Roberto, Rasheed Tinku, Narayanan Rajesh. Virtual network functions orchestration in enterprise WLANs. In: Proc. of IEEE ManFI; 2015; Ottawa, Canada.

33. Mijumbi R., Serrat J., Gorricho J. L., Bouten N., Turck F. De, Davy S.. Design and evaluation of algorithms for mapping and scheduling of virtual network functions. In: in Proc. of IEEE NetSoft; 2015; Seoul, Korea.
34. *Further study on critical C-RAN technologies, VI.0.* : Next Generation Mobile Networks Alliance Frankfurt, Germany; 2015.
35. Jinri Huang, Yannan Yuan. White Paper of next Generation Fronthaul Interface. 2015;.
36. Rost P., Bernardos C. J., Domenico A. D., et al. Cloud technologies for flexible 5G radio access networks. *IEEE Communications Magazine*. 2014;52(5):68-76.
37. Wubben Dirk, Rost Peter, Bartelt Jens Steven, et al. Benefits and impact of cloud computing on 5G signal processing: Flexible centralization through cloud-RAN. *Signal Processing Magazine, IEEE*. 2014;31(6):35–44.
38. Checko A., Avramova A. P., Berger M. S., Christiansen H. L.. Evaluating C-RAN fronthaul functional splits in terms of network level energy and cost savings. *Journal of Communications and Networks*. 2016;18(2):162-172.
39. Arnold Paul, Bayer Nico, Belschner Jakob, Zimmermann Gerd. 5G radio access network architecture based on flexible functional control/user plane splits. In: Proc. of EuCNC; 2017; Oulu, Finland.
40. Liu Jingchu, Zhou Sheng, Gong Jie, Niu Zhisheng, Xu Shugong. Graph-based framework for flexible baseband function splitting and placement in C-RAN. In: Proc. of IEEE ICC; 2015; London, United Kingdom.
41. Maeder Andreas, Lalam Massinissa, De Domenico Antonio, et al. Towards a flexible functional split for Cloud-RAN networks. In: Proc. of EuCNC; 2014; Bologna, Italy.
42. Dötsch Uwe, Doll Mark, Mayer Hans-Peter, Schaich Frank, Segel Jonathan, Sehier Philippe. Quantitative analysis of split base station processing and determination of advantageous architectures for LTE. *Bell Labs Technical Journal*. 2013;18(1):105–128.
43. Bartelt Jens, Rost Peter, Wubben Dirk, Lessmann Johannes, Melis Bruno, Fettweis Gerhard. Fronthaul and backhaul requirements of flexibly centralized radio access networks. *Wireless Communications, IEEE*. 2015;22(5):105–111.
44. Harutyunyan Davit, Riggio Roberto. Flex5G: Flexible Functional Split in 5G Networks. *IEEE Transactions on Network and Service Management*. 2018;15(3):961–975.
45. Koutsopoulos Iordanis. Optimal functional split selection and scheduling policies in 5G Radio Access Networks. In: Proc. of IEEE ICC Workshop; 2017; Paris, France.
46. Checko Aleksandra, Holm Henrik, Christiansen Henrik. Optimizing small cell deployment by the use of C-RANs. In: Proc. of European Wireless; 2014; Barcelona, Spain.
47. Fischer Anath, Botero Juan Felipe, Till Beck Michael, De Meer Hermann, Hesselbach Xavier. Virtual network embedding: A survey. *Communications Surveys Tutorials, IEEE*. 2013;15(4):1888–1906.
48. Chowdhury Mosharaf, Rahman Muntasir Raihan, Boutaba Raouf. Vineyard: Virtual network embedding algorithms with coordinated node and link mapping. *IEEE/ACM Transactions on Networking (TON)*. 2012;20(1):206–219.
49. Zhang Sheng, Qian Zhuzhong, Wu Jie, Lu Sanglu, Epstein Leah. Virtual network embedding with opportunistic resource sharing. *IEEE Transactions on Parallel and Distributed Systems*. 2014;25(3):816–827.
50. Rappaport T. S., Sun S., Mayzus R., et al. Millimeter Wave Mobile Communications for 5G Cellular: It Will Work!. *IEEE Access*. 2013;1:335-349.
51. Pi Z., Khan F.. An introduction to millimeter-wave mobile broadband systems. *IEEE Communications Magazine*. 2011;49(6):101-107.
52. Ghosh A., Thomas T. A., Cudak M. C., et al. Millimeter-Wave Enhanced Local Area Systems: A High-Data-Rate Approach for Future Wireless Networks. *IEEE Journal on Selected Areas in Communications*. 2014;32(6):1152-1163.

53. Rangan Sundeep, Rappaport Theodore S, Erkip Elza. Millimeter-wave cellular wireless networks: Potentials and challenges. *Proceedings of the IEEE*. 2014;102(3):366–385.



TABLE 1 Fronthaul bandwidth and one-way latency requirements (absolute and relative) for the functional splits considered in this work⁹.

Functional splits	DL bandwidth	Latency	Latency class
PHY–RF Split	2.46 Gbps ($\times 1$)	250 μ s ($\times 1$)	Ideal
PHY Split	0.93 Gbps ($\times 2.5$)	2 ms ($\times 8$)	Near Ideal
MAC Split	0.15 Gbps ($\times 16.5$)	6 ms ($\times 24$)	Sub Ideal
PDCP–RLC Split	0.15 Gbps ($\times 16.5$)	30 ms ($\times 120$)	Non Ideal

TABLE 2 Terminologies of RAN components and functional splits by different organizations.

RAN Components and Functional Splits			
SCF ⁹	3GPP ⁴	NGFI ³⁵	NGMN ³⁴
BBU (Baseband Unit)	CU (Centralized Unit)	RCC (Radio Cloud Center)	DU (Digital Unit)
RRH (Remote Radio Head)	DU (Distributed Unit)	RRU (Remote Radio Unit)	RU (Radio Unit)
C-RAN	PHY-RF split (option 8)	Baseband / RF function split	C-RAN
PHY split	Intra PHY split (option 7)	Symbol level / Sample level split	-
MAC split	Intra MAC split (option 5)	High MAC / Low MAC split	-
PDCP–RLC split	PDCP–high RLC split (option 2)	-	-

TABLE 3 Substrate network parameters

Variable	Description
G_s	Substrate network graph.
N_s	Set of substrate nodes in G_s .
N_s^{du}	Set of substrate DUs in G_s .
N_s^{cu}	Set of substrate CU pools in G_s .
E_s	Set of substrate links in G_s .
$\omega_a^s(n)$	Number of RF front–ends available at the node $n \in N_s$.
$\omega_i^s(n)$	Set of MMW interfaces available at the node $n \in N_s$.
$\omega_c^s(n)$	The processing capacity of the node $n \in N_s$.
$\omega_b^s(e^{nm})$	Capacity of the MMW link $e^{nm} \in E_s$ (in Gbps).
$\delta(n)$	Coverage radius of the DU $n \in N_s^{du}$ (in meters).
$loc(n)$	Geographical location of the node $n \in N_s$ (x, y).
Λ_i	Cost for using the MMW interface $i \in \omega_i^s(n)$ of the node $n \in N_s$.
Λ_e	Cost for each unit of bandwidth resources of the substrate link $e \in E_s$.

TABLE 4 Virtual network request parameters

Variable	Description
G_v	Virtual network graph.
N_v	Set of virtual nodes in G_v .
N_v^{du}	Set of virtual DUs in G_v .
N_v^{cu}	Set of virtual CU pools in G_v .
E_v	Set of virtual links in G_v .
$\omega_a^v(n)$	Number of RF front–ends required by the node $n \in N_v$.
$\omega_c^v(n)$	The processing requirement of the node $n \in N_v$.
$\omega_b^v(e^{nm})$	Requested capacity for the CPRI link $e^{nm} \in E_v$ (in Gbps).
$\omega_e^v(e^{nm})$	Virtual link embedding cost $e^{nm} \in E_v$.
$loc(n)$	Desired geographical location for the DU $n \in N_v^{du}$ (x, y).

TABLE 5 CPRI link bandwidth per option.

CPRI option	CPRI rate	I/Q sampling data rate	LTE configurations
1	600 Mbps	400 Mbps	10 MHz, 1x1SISO
2	1.2 Gbps	0.9 Gbps	20 MHz, 1x1SISO
3	2.5 Gbps	1.8 Gbps	20 MHz, 2x2MIMO
5	5 Gbps	3.6 Gbps	20 MHz, 4x4MIMO

TABLE 6 Fronthaul bandwidth and processing requirements for the DU and the CU pool of the small cell n for the functional splits considered in this work.

Splits	Resources	Processing capacity		DL bandwidth	
		DU	CU pool	1x1SISO	2x2MIMO
PHY–RF Split		$0 \cdot \omega_c^v(n)$	$1 \cdot \omega_c^v(n)$	1.23 Gbps	2.46 Gbps
PHY Split		$0.5 \cdot \omega_c^v(n)$	$0.5 \cdot \omega_c^v(n)$	0.46 Gbps	0.93 Gbps
MAC Split		$0.7 \cdot \omega_c^v(n)$	$0.3 \cdot \omega_c^v(n)$	0.07 Gbps	0.15 Gbps
PDCP–RLC Split		$0.9 \cdot \omega_c^v(n)$	$0.1 \cdot \omega_c^v(n)$	0.07 Gbps	0.15 Gbps

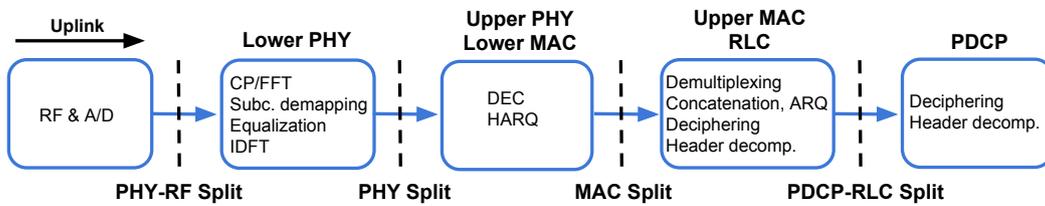


FIGURE 1 Signal processing along with some of the functional split options within the RAN protocol stack in LTE networks.

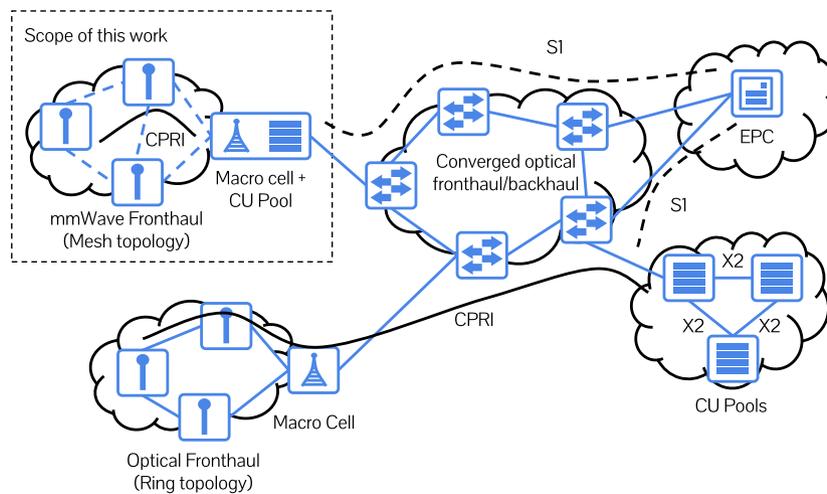


FIGURE 2 The network architecture used for this CU placement problem. The black solid lines represent the CPRI links while the black dashed lines represent the S1 links. CU pools and the macro cells are co-located, therefore, improving computational locality and shortening the fiber-based CPRI links.

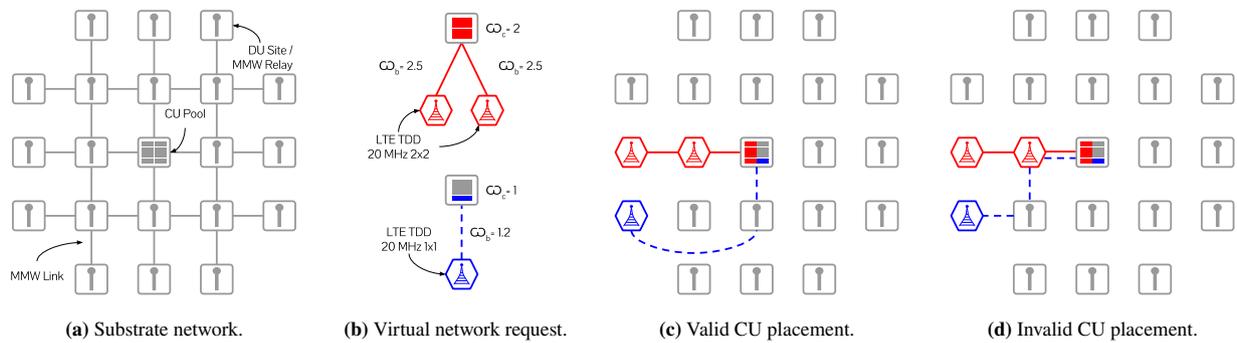


FIGURE 3 Sample substrate network, virtual network request and CU placement. Notice how the CU placement in Fig.3c is valid since the interface constraint is satisfied on all relay DUs. Conversely, the CU placement in Fig.3d is invalid since it would require four interfaces in relaying DUs while only two are actually available.

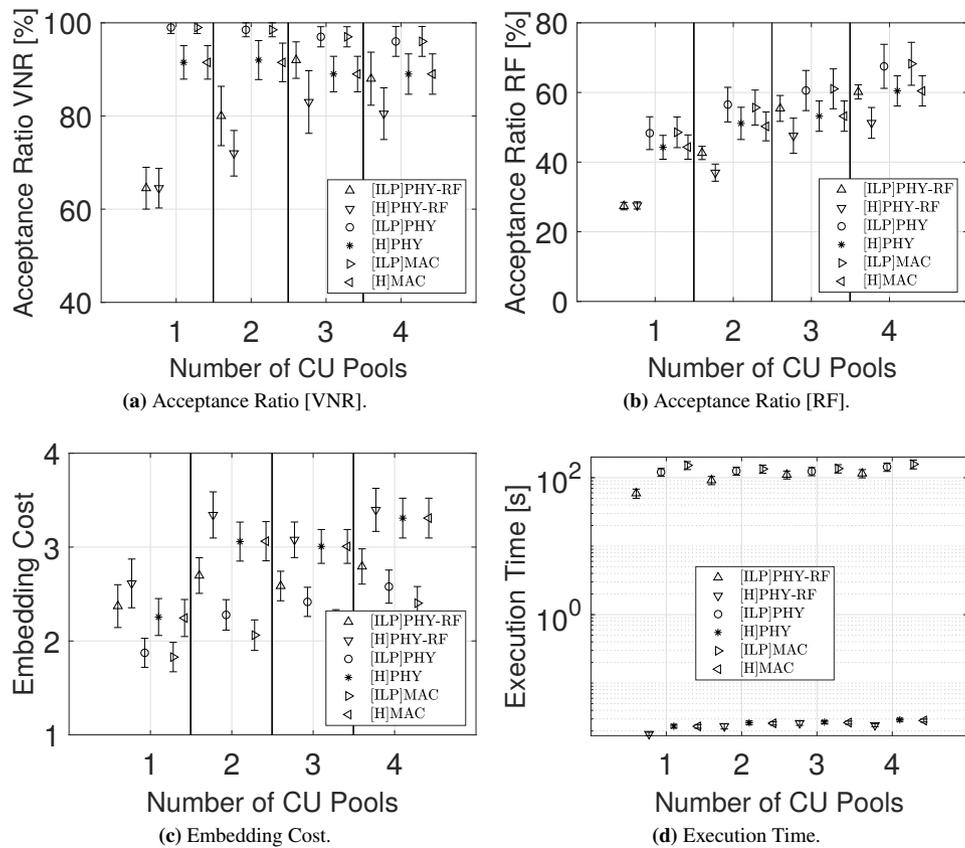


FIGURE 4 Performance of the ILP-based algorithm and of the heuristic with a different number of substrate CU pools for the considered functional splits.

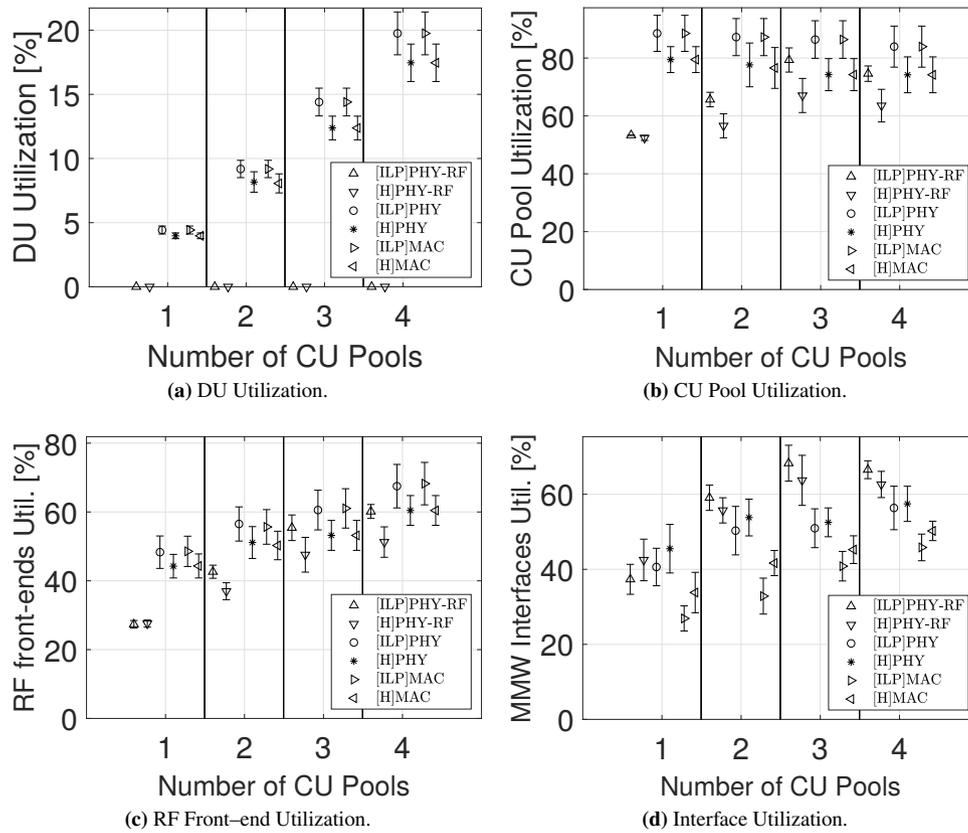


FIGURE 5 Substrate resource utilizations of the ILP-based algorithm and of the heuristic.

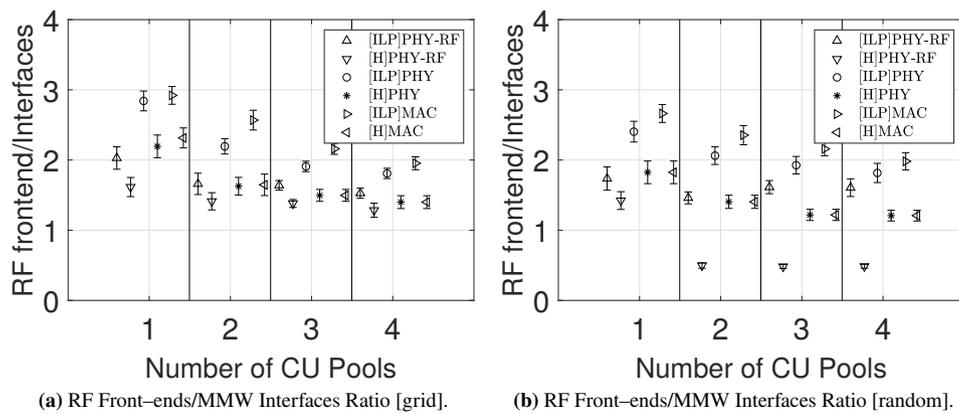


FIGURE 6 The utilization ratio of the RF front-ends and MMW interfaces of the ILP-based placement algorithm and the placement heuristic for the grid and the random networks with a different number of substrate CU pools.

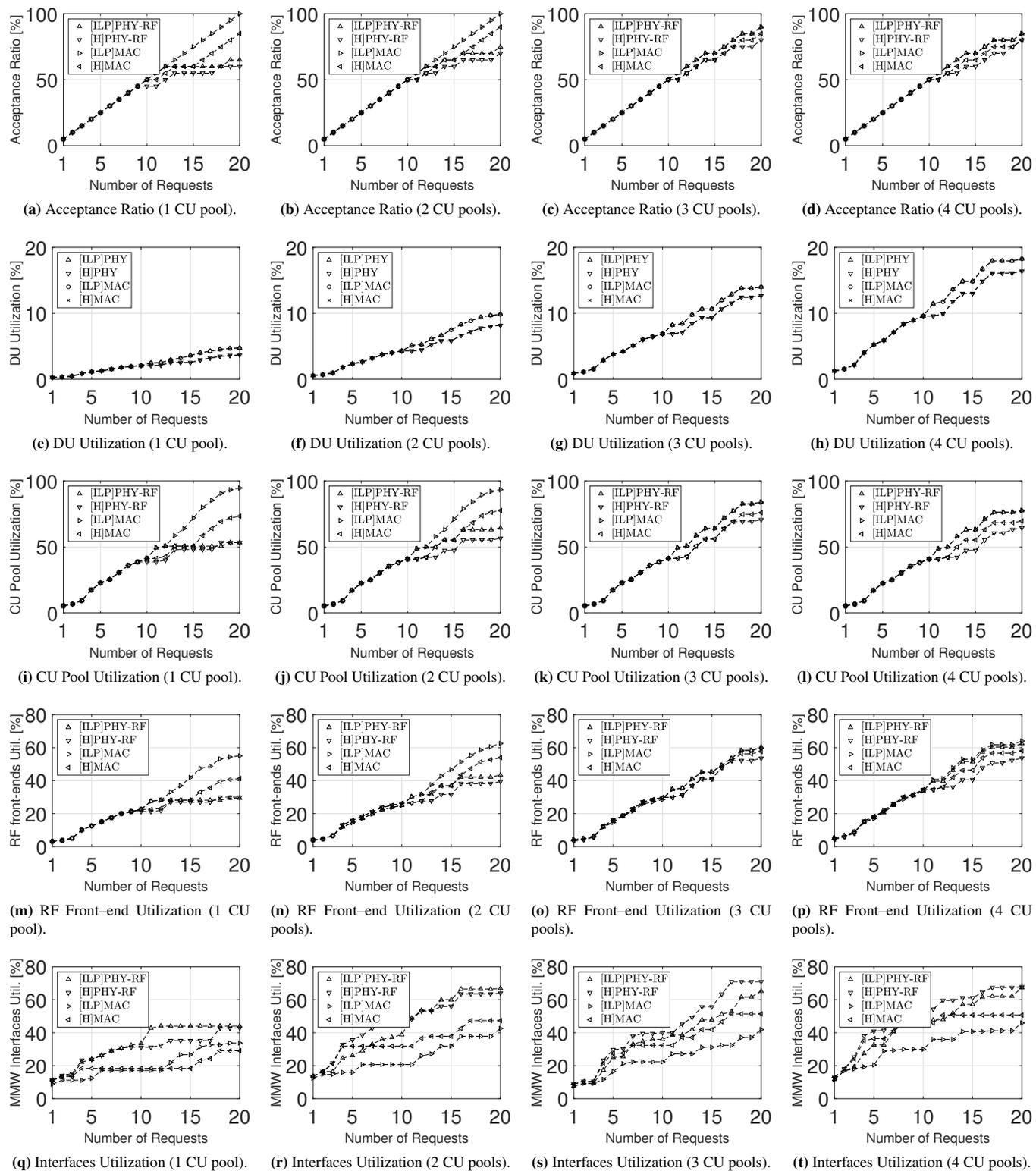


FIGURE 7 Performance of the ILP-based placement algorithm and of the placement heuristic with a different number of substrate CU pools for one simulation (20 embeddings) for the grid-shaped substrate network.