# Dynamic DU/CU Placement for 3-layer C-RANs in Optical Metro-Access Networks

**Leila Askari, Francesco Musumeci, Lorenzo Salerno, Omran Ayoub, Massimo Tornatore**
*Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy*
*email: firstname.lastname@polimi.it*

**ABSTRACT**
We investigate power-aware dynamic DU/CU placement for a 3-layer C-RAN architecture in optical metro-access networks. We show that adaptive placement and reconfiguration of both DU and CU locations based on traffic allows to significantly decrease energy consumption, paying off only a negligible increase in service blocking.
**Keywords**: RAN, DU/CU placement

## 1. INTRODUCTION

3rd Generation Partnership Project (3GPP) has recently proposed a new architecture for Centralized Radio Access Networks (C-RAN) based on separating baseband functions into three entities or 'splits' [1]. These three splits are the Centralized Unit (CU), the Distributed Unit (DU), and Radio Unit (RU). CUs are usually placed in higher-hierarchy metro nodes, e.g., a large central office (CO), and are responsible for higher-layer processing, while the DUs, which are responsible for lower-layer processing, are distributed nearer to antennas/RUs [2].

With respect to the existing 2-layer architecture (where processing functions are distributed only between two elements, RRH and BBU [3]), the 3-layer architecture allows a more flexible deployment of baseband functions. Moreover, it is expected that an intelligent placement of CUs and DUs can yield further advantages in terms of cost, power consumption and service blocking [4]. How to reach these objectives while meeting service latency requirements is not trivial. For instance, centralizing the baseband functions reduces energy consumption, but increases substantially the required transport capacity. On the contrary, a more distributed placement of CUs and DUs allow decreasing network-capacity requirements but yields excessive energy consumption. It is therefore decisive to leverage the flexibility offered by the 3-layer architecture to jointly meet capacity, latency and energy requirements. To this end, we propose a novel algorithm that decides the placement of CUs and DUs and performs routing, grooming and wavelength assignment, and we compare its performance to that of baseline strategies in terms of power consumption and blocking probability in optical metro-access networks.

The DU/CU placement problem has been already investigated in recent studies, however most of them have looked at the static version of the problem. Ref. [5] modeled the CU placement problem through an Integer Linear Program (ILP) and considered different split options with the objective of reducing network power consumption. Similarly, Ref. [6] proposed an ILP to minimize the computational capacity required to support 5G network functions, while satisfying delay constraints of chosen splits. Our work focuses on the power-minimized DU/CU placement for dynamic traffic considering the 3-layer architecture for C-RAN. To the best of our knowledge, this is the first work tackling this problem in literature. To address this problem, we provide a power model to estimate power consumption due to traffic processing at DUs and CUs, and propose an algorithm for energy-efficient placement of CUs and DUs and traffic routing and wavelength assignment in dynamic optical metro-access networks.

Figure 1 shows the different functional split options to define the bandwidth and latency requirement for different parts of the network. In other words, using this functional separation, the 2-layer C-RAN is now transformed to a 3-layer architecture consisting of *fronthaul*, used to transport traffic between RUs and DUs, *midhaul* that connects DUs to CUs, and *backhaul* connecting CUs to the core network. The baseband functions are as follows: Radio Resource Control (RRC) that is responsible for handover and managing security functions, Packet Data Convergence Protocol (PDCP), that is responsible for security functions such as cyphering, Radio Link Control (RLC) that is responsible to perform segmentation and reassembly for higher layers, Media Access Control (MAC) that is responsible for multiplexing data from different radio bearers, Physical layer (PHY) that converts the radio signal to digital bits for downlink and vice versa for uplink and Radio Frequency (RF). The interface between CU and DU is *F1,* while the interface between DU and RU is *Fx*. In our work, we consider option 2 and option 7 as they have been identified as the main standard split options by ITU [7].

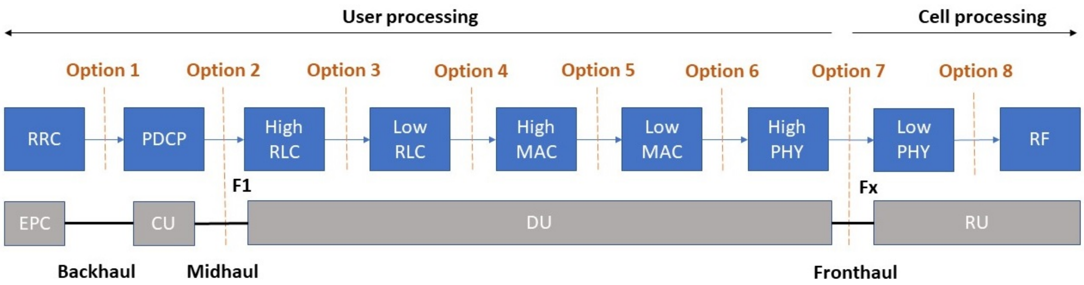


*Figure 1. Functional split*

## 2. Power-aware DU/CU placement and traffic routing and wavelength assignment

The problem addressed in this paper can be formally stated as follows. Given a metro-access network, where a subset of nodes hosts a hierarchy of Central Offices (COs) enhanced with computational resources to host CU/DUs (hotels), and given dynamic traffic demands originated by cell sites and directed towards Core CO, we decide the placement of DUs and CUs and perform routing, grooming and wavelength assignment of traffic demands. Our objective to minimize power consumption is achieved by minimizing number of active CU/DU hotels while satisfying constraints on link capacity and maximum fronthaul latency.

### 2.1 Power model

Power consumption of DU and CU can be divided into three main components: processing card, interface and common site infrastructure (CSI). To calculate power consumption of processing card, we need first to estimate the complexity of functions inside DU and CU and its dependency on the amount of processed traffic. We used the model in [8] which is based on the complexity values of each function calculated based on Giga Operations Per Second (GOPS). Power consumption of a function $i$ is calculated as:

$$P_i = \frac{C_i}{T} \ \Gamma i \qquad (1)$$

where $C_i$ is complexity value of this baseband function $i$ and $T$ is a technology-dependent factor that indicates hardware complexity based on system configuration and on the date of system deployment. Note that, in Eqn. 1, $\Gamma_i = 1$ if we consider a reference scenario in which the RU operates at 20 MHz bandwidth, fully loaded system, single-input-single-output antenna configuration, spectral efficiency 6bps/Hz, and 64-QAM modulation scheme with coding rate 1. In case of different radio-configuration scenarios, a different scaling factor $\Gamma_i$ must be used to calculate corresponding $P_i$. This scaling factor depends on various input parameters, which we define through set $X$=:{bandwidth, spectral efficiency, number of antennas, system load, number of spatial streams, quantization (e.g., 4bits, 16bits and 24bits)}. The impact of each parameter is defined by scaling exponent $S_{i,x}$ for each baseband function $i$ and parameter $x$ in $X$. To calculate the value of $\Gamma_i$ we use the following formula:

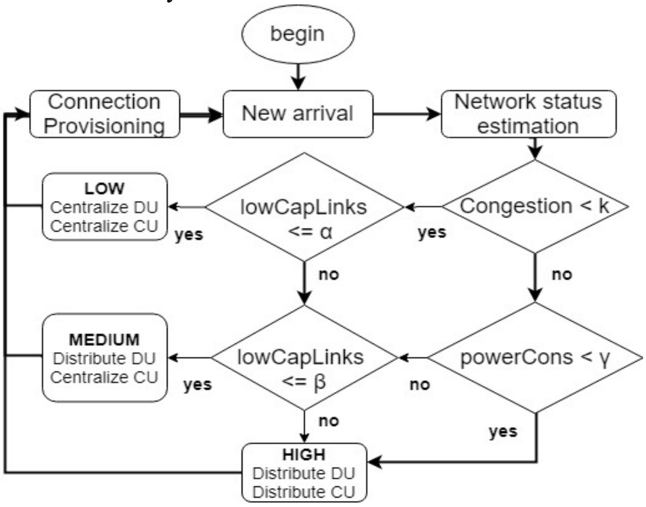


*Figure 2. Flowchart of DDUP algorithm*

$$\Gamma_i = \prod_{x \in X} \left(\frac{x_{act}}{x_{ref}}\right)^{S_{i,x}} \qquad (2)$$

in which $x_{act}$ is the actual value of the parameter $x$, and $x_{ref}$ is the value of parameter $x$ for reference scenario. For each function $i$ and parameter $x$, the scaling exponent $S_{i,x}$ is taken from [8]. Therefore, based on the functions placed in DU and CU, we can calculate $P_{DU}$ and $P_{CU,}$ respectively. Let us now analyse our modelling of the power consumption for the interfaces: *i*) for fronthaul ($P_{FH}$), we considered the value 18.2 W as in [9]; *ii*) for backhaul ($P_{BH}$), we considered the value of 1 W as in [8]; *iii*) midhaul ($P_{MH}$) we considered a constant value of 10W.

The power consumption of CSI, $P_{CSI}$, accounts for cooling, lights, and for the monitoring system. Therefore, in a first approximation, $P_{CSI}$ is not dependent on the number of CUs/DUs in a hotel node and is assumed to be constant. We used the value 2100 W as in [8] for this parameter. Now that we have the power consumption of all their components, we can calculate the power consumption of DU and CU as follows:

$$P_{DU-Hotel} = N_{DU}\, P_{DU} + P_{FH} + P_{MH} + P_{CSI} \quad (3) \qquad P_{CU-Hotel} = N_{CU}\, P_{CU} + P_{MH} + P_{BH} + P_{CSI} \quad (4)$$

where $N_{DU}$ is the number of DUs in the DU-*hotel* node and $N_{CU}$ is the number of CUs in the *CU-hotel* node. As for idle hotel nodes, we considered a fixed power consumption of 300 W [10].

### 2.2 Dynamic DU/CU placement algorithm

The high-level flowchart of the proposed Dynamic DU/CU Placement (DDUP) algorithm to perform DU/CU placement in a metro network is depicted in Fig. 2. As a first step, DDUP collects statistics on the utilization of network links (network status). Specifically, link utilization of all links is checked against a pre-defined threshold, which determines link congestion. If the number of congested links in the network is below $k$ percent, and the number of links that have capacity for just one more fronthaul connection (*lowCapacityLinks* in flowchart) is below a certain percentage threshold α, DDUP tries to implement *LOW* policy, that is, to centralize CUs and DUs by putting them both in the Core CO. To this end, algorithm tries to place at first the DU in the Core CO or any other closest node to the Core CO. Upon successful placement of DU, CU is placed either in the Core CO or the closest node to the Core CO. Note that the maximum tolerated latency for fronthaul, midhaul and backhaul should always be satisfied. Conversely, if more than $k$ percent of links are congested and power consumption of hotel nodes, calculated using Enqs. 3 and 4, is less than γ, but *lowCapacityLinks* is higher than $\beta$, DDUP tries to distribute DUs as much as possible. To do so, it chooses an already active DU-hotel node that satisfies the fronthau latency requirement. If such node is not found, DDUP activates a new DU-hotel and places the DU on that node. After that, the *LOW* policy is followed to place CU. This policy is referred to as *MEDIUM*. If links are highly congested i.e., percentage of links that do not have enough capacity for a new fronthaul connection is more than $\beta$, at first DDUP tries to follow the same policy as *MEDIUM* to place DU. After successful placement of DU the same policy is followed to find a suitable CU-hotel and place CU. In other words, the *HIGH* policy tries to distribute both DUs and CUs in the network. After deciding about placement of DU and CU, grooming, routing and wavelength assignment are performed as described in our previous work [4]. Note that, adjustment of all the thresholds can be decided based on the policies of the network operator in terms of power consumption and number of served connections. That is, defining a very high threshold for, e.g., $k$, results in a smaller number of served connections, while defining a low value of $k$ results in more activated hotels, and therefore higher power consumption.

## 3. NUMERICAL RESULTS

We perform analysis using a discrete-event simulator we developed in C++. We randomly generate connection arrivals originating from RUs according to a truncated Poisson distribution with mean holding time of μ=1 (each connection requiring a CU and a DU instance). We run the simulations for 50000 demands, such that all plotted results are within 5% confidence interval with 95% confidence level. We consider the realistic metro topology in Fig. 3, with 60 Wavelength-Division-Multiplexing (WDM) links and 51 nodes (12 of which are hotel nodes, i.e., they are equipped with computational capacity, and can host up to 2000 CUs/DUs instances each). Each WDM link supports 16 wavelengths at 100 Gbps. All other parameters in our evaluations are summarized in *Table 1.*

*Table 1. Parameters used for numerical evaluation.*

| Parameter | k | α | β | γ | Required fronthaul rate | Required backhaul rate | Max fronthaul, midhaul, backhaul latency |
|---|---|---|---|---|---|---|---|
| Value | 30% | 20% | 40% | 30kW | 10 Gbit/s | 300 Mbit/s | 100μs, 1ms, 40ms |

### 3.1 Benchmark algorithms

We developed three benchmark algorithms to be compared with DDUP:

- *Fully centralized* places both DUs and CUs in the Core CO or as close as possible to it.
- *Fully distributed* tries to place DUs and CUs at the edge of the network, as close as possible to the RUs.
- *Mixed* places DUs as close as possible to the RUs while CUs are placed in the Core CO.

### 3.2 Performance comparison

We evaluated the performance of our algorithm considering the following metrics: i) average number of active hotels, calculated as the number of nodes having at least one CU/DU active on them weighted by amount of time this node is serving demand requests; ii) hotels power consumption, calculated based on Eqns. (3) and (4) respectively, for active DU hotels and CU hotels iii) blocking probability, calculated as the number of blocked demands over the total number of requested demands during the simulation.

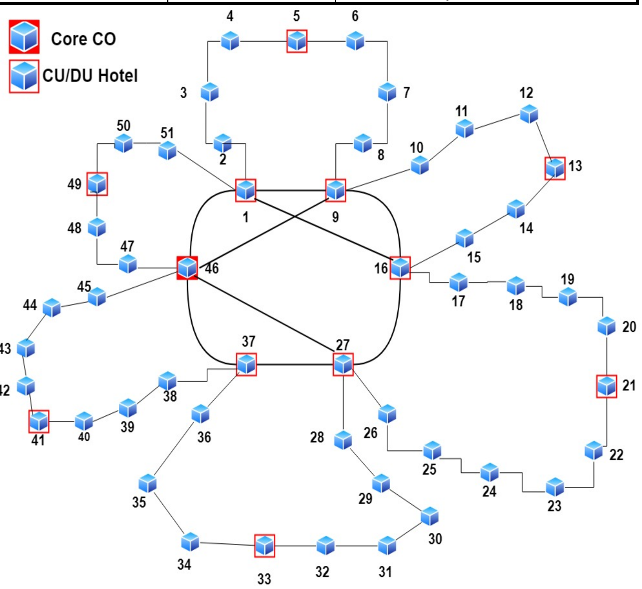

*Figure 3. WDM metro network topology*

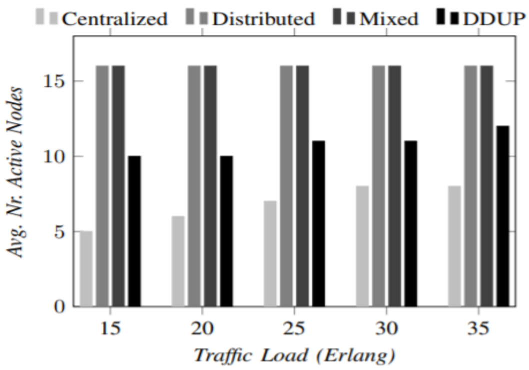


*Figure 4. Average number of active hotels.*

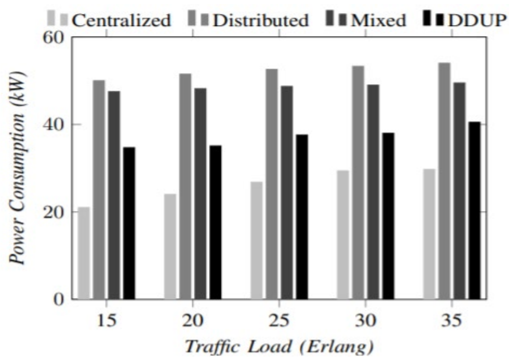


*Figure 5. Power consumption of hotels.*

Note that blocking can happen either due to lack of link bandwidth or due to violation of latency requirements. Average number of active hotels is depicted in Fig. 4. As expected, the *Fully centralized* approach activates a smaller number of hotels with respect to other approaches, since the objective of this approach is to place DU and CUs as much as possible in Core CO. Instead, both *Mixed* and *Fully distributed* activate all the 12 hotel nodes, as these two approaches distribute DUs in the network. The number of active hotels for DDUP always lies between the values for other algorithms and for the lower traffic loads our approach activates almost half of the hotel nodes. Figure 5 shows the power consumption of hotel nodes that depends on the number of active CUs/DUs in a hotel node. For lower traffic loads, DDUP has a power consumption up to 27% less with respect to *Mixed* and *Fully distributed* approaches, as DDUP tries to activate DUs and CUs in as few hotel nodes as possible by re-using already active ones.

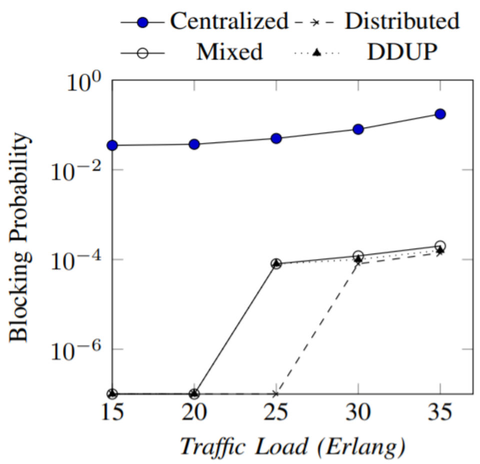


*Figure 6. Blocking probability*

Finally, Fig. 6 compares blocking probability of the different approaches. Blocking probability for DDUP always lies between the upper bound (*Fully Centralized*) and lower bound (*Fully distributed*). Note that, for higher traffic loads the blocking probability for our approach reaches a value very close to *Mixed* approach even if the power consumption of our approach is significantly lower. This is due to the fact that, when network congestion reaches a pre-defined threshold, DDUP starts to distribute CUs and DUs as much as possible in the network, closer to the RUs. However, for the lower traffic loads, since the links in the network are not congested, our approaches tries to centralize as much as possible to decrease the power consumption.

## 4. CONCLUSIONS

We proposed an algorithm to place DUs and CUs in an optical metro network with the aim to reduce power consumption by consolidating CUs and DUs in less hotel nodes, while considering the latency requirements and hotel nodes capacity. Numerical results show our approach can achieve up to 33% reduction in power consumption, while maintaining an acceptable blocking probability compared to baseline approaches.

## ACKNOWLEDGEMENTS

This work was supported by the European Community under grant agreement no.761727 Metro-Haul project.

## REFERENCES

[1] 3GPP TR 38.801 V14.0.0 (2017-03), Radio access architecture and interfaces (Release 14)
[2] L. M. P. Larsen, et al., in IEEE Communications Surveys & Tutorials, vol. 21, no. 1, pp. 146-172, 2019.
[3] https://www.gsma.com/futurenetworks/wiki/5g-era-mobile-network-cost-evolution/
[4] F. Musumeci, et al., in Journal of Optical Communications and Networking, vol. 11, no. 4, pp. B71--B82, 2019.
[5] L. M. M. Zorello, et al., in Optical Network Design and Modelling conference, 2020.
[6] A. N. Al-Quzweeni, et al., in IEEE Access, vol. 7, pp. 44939-44958, 2019.
[7] G.sup.5GP: 5G Wireless Fronthaul Requirements in a PON Context, ITU-T, 2020.
[8] B. Debaillie, et al., in IEEE 81st Vehicular Technology Conference, 2015.
[9] COMBO Project: https://www.ict-combo.eu.
[10] M.Khan, et al., in European Conference on Network and Communications, 2015.