DU/CU Placement for C-RAN over Optical Metro-Aggregation Networks

Hao Yu, Francesco Musumeci, Jiawei Zhang, Yuming Xiao, Massimo Tornatore, Yuefeng Ji

DU/CU Placement for C-RAN over Optical Metro-Aggregation Networks

Hao Yu[1,2][0000−0003−3522−3426], Francesco Musumeci[2], Jiawei Zhang[1], Yuming Xiao[1], Massimo Tornatore[1], and Yuefeng Ji[2]

[1] Beijing University of Posts and Telecommunications, Beijing, China
[2] Politecnico Di Milano, Milan, Italy
{yuhao92, zjw, yumingxiao, jyf}@bupt.edu.cn {francesco.musumeci, massimo.tornatore}@polimi.it

Abstract. To meet emerging mobile traffic requirements, Centralized Radio Access Network (C-RAN) has been proposed to split the base station (BS) into two functional entities: the baseband units (BBU) and the remote radio heads (RRH). In C-RAN, by centralizing BBUs into BBU pools and leaving the RRHs in the cell sites, significant cost and energy savings and improved radio coordination can be achieved. However, C-RAN requires a costly high-capacity and low-latency access/aggregation network to support fronthaul traffic (i.e., digitized baseband signal). Hence, more recently, a new C-RAN architecture has been proposed (i.e., by 3GPP, IEEE 1914 WG), that defines three baseband function entities (or splits): central unit (CU), distributed unit (DU) and remote unit (RU). These three entities are expected to be interconnected by two external interfaces, called F1 and Fx. By transforming the RAN into a 3-layer (CU-DU-RU) architecture, more flexible deployment of the baseband functions can be achieved that better adapts to the heterogeneous characteristics of incoming 5G service requirements. It is also expected that, by properly placing CUs and DUs in the metro/aggregation network, higher benefits in terms of cost and power consumption can be achieved with respect to the previous 2-layer (BBU-RRH) architecture. In this paper, we investigate the optimal CU/DU placement problem in a 3-layer RAN architecture and formalize it by integer linear programming. We evaluate the benefits of the 3-layer architecture compared to the 2-layer architecture, showing that the consolidation degree of baseband processing depends heavily on fronthaul traffic latency, transport network capacity and processing capacity.

Keywords: C-RAN · functional splits · fronthaul · baseband processing deployment · wavelength division multiplexing · 5G

1 Introduction

Ever-increasing mobile-traffic demand requires operators to deploy more base stations and to keep updating their radio access network (RAN). In the future, 5G RAN is set to provide an even larger variety of services with widely varying requirements on bandwidth and latency. Therefore, future RAN design is

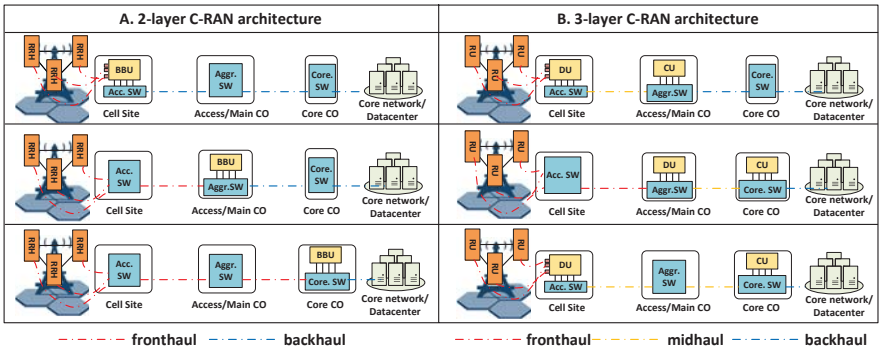


Fig. 1: Illustration of 2-layer and 3-layer C-RAN architecture

expected to satisfy the very stringent constraints in terms of tolerable latency and required data rate [1], e.g., end-to-end latency requirements below 1 ms, more than 250 Gb/s/km^2 in dense-urban areas and device density in the order of thousands per km^2 [2]. Overall, a very large amount of baseband processing resources and network resources will be required to be deployed in the RAN.

In 4-th generation LTE networks, Distributed Radio Access Network (D-RAN) is the dominant RAN paradigm, where the Base Station (BS) comprises two modules: Remote Radio Head (RRH) for transmission and reception of radio signals, frequency up/down conversion and power amplification, and Baseband Unit (BBU) to perform digital processing of baseband signal. The BBU and RRH were traditionally co-located in the same housing facility, so the maintenance and investment costs of networks increase linearly with the number of BSs, which leads to a not scalable solution. An alternative RAN architecture, called Centralized Radio Access Network (C-RAN), has been proposed as a scalable solution in terms of both power and cost efficiency. In C-RAN, the BBUs are centralized into larger housing facilities, called BBU pools, which are connected with the RRHs through a high-capacity and low-latency "fronthaul" network [3]. Optical metro/aggregation networks based on wavelength division multiplexing (WDM) are considered as a promising candidate for transport network of RAN. Besides performing traffic aggregation and switching (SW), the central offices (COs) at different hierarchical stages of the metro network can be also equipped with processing resources to process baseband signal. The procedure of moving BBUs from cell sites to COs is called BBU hotelling, as shown in Fig. 1(left). From cell site to access/main CO, to core CO, the fewer COs need to be used as BBU hotels, and the more "consolidation" is achieved through BBU hotelling. As the motivations of BBU hotelling are to allow operators to save costly installation and maintenance of processing facilities inside COs, centralizing more BBU into COs in higher stages of networks can increase these benefits.

However, due to the high fronthaul cost of C-RAN [4], RAN architecture is further evolving towards decentralizing part of baseband processing functions at

the network edge, in order to decrease the fronthaul bandwidth. BBU is now divided into two parts: Distributed Unit (DU) and Central Unit (CU). The DU contains some real-time baseband processing functions, i.e., Hybrid Automatic Repeat Request (HARQ) processing, radio coordination, while CU contains some non real-time baseband processing functions. Original RRH with part of PHY-layer functions comprises Remote Unit (RU). It follows that the RAN architecture evolves from a 2-layer (BBU-RRH) to 3-layer (CU-DU-RU) architecture. Through flexible distribution of CUs and DUs in the metro transport network, service requirements can be satisfied according to different DU/CU hotelling schemes. As shown in the right side of Fig. 1, the RU replaces the RRH at the cell sites, and DUs/CUs can be placed at any locations starting from the cell sites up to the COs at the different stages of the metro network. Compared with 2-layer RAN architecture, a new network segment is introduced, called midhaul, which connects DU and CU. In this paper, we re-consider the DU/CU placement problem in the 3-layer RAN architecture by minimizing the number of active COs for CU/DU hotelling. We also provide a mathematical model for this placement problem, and we apply this model over a limited, yet realistic network scenario. This model allows us to show the interplay between DU/CU placement and front/mid/back-haul, and to investigate the relation between the consolidation of baseband processing functions and the processing/bandwidth capacity constraints, as well as service latency. For the metro transport solutions, we consider OTN (more specifically, we assume a version of OTN optimized for mobile transport, called M3C-OTN, aiming at Mobile optimized, Multi-service and Metro Cloud OTN solution for the 5G scenario) and overlay, in line with the assumptions used in [8].

The remainder of this paper is organized as follows. Section II introduces the related works. Section III introduces functional split based 3-layer RAN architecture. Section IV illustrates joint DU/CU placement problem in the metro networks. Section V shows the illustrative numerical results. Section VI concludes the work.

2 Related Works

The problem of baseband functions placement and traffic routing has been investigated in recent years. In traditional C-RAN architecture, authors in [5] propose the energy-efficient virtual base station (vBS) formation problem, where they combine the virtual passive optical network (vPON) with vBS to form a virtual RAN (vRAN) and minimize the number of active digital units (DUs) in the DU pool and the number of wavelengths in the PON at the same time. In [6], a "BBU aggregation" problem is proposed to minimize the number of the active BBUs for energy saving. BBU aggregation is modeled as an evolved 2D bin-packing problem and two heuristics to solve this problem are proposed. Authors in [7] motivate energy-efficient BBU aggregation problem in an AWGR-based passive WDM network. They introduce the AWGR decomposition into the BBU aggregation to help reduce the cost of the tunable transceivers and BBUs.

The problem of BBU placement over a metro/aggregation network was first proposed in [8], where the relation between latency and the consolidation of BBUs is discussed under the OTN and overlay cases separately. In addition to BBU placement problem, BBU pool allocation and selection problem for C-RAN is also proposed in [9], the authors investigated how to deploy BBU pool among the optical network nodes and how to choose BBU pool to host the BBU of each traffic request with the objective of maximizing traffic acceptance ratio and minimizing network resource usage.

For baseband function placement in a functional spilt RAN architecture, ref. [10] proposes a graph-based framework for flexible baseband function splitting and placement problem, and determines how to split the baseband function chain for each cell to maximize the utilization of processing resources under the constraint of latency and processing capacity using genetic algorithm. In [11], the authors propose a fully flexible functional split RAN architecture and define a new baseband entity, called flexible unit (FU). Based on this, they motivate a minimal number of active central office problem through the energy-efficient placement of FUs.

In summary, most existing works consider baseband function placement in the 2-layer RAN, while there are no works providing the analysis of the CU and DU placement problem in 3-layer C-RAN based on the 3GPP standard functional split options and the relation between the consolidation of CU/DU and constraints of latency, network capacity, as the one provided in this paper.

3 Functional Split based 3-layer C-RAN Architecture

According to 3GPP [12] and other standardization bodies, the 5G RAN architecture has been defined as a 3-layer architecture consisting of a CU, a DU, and a RU. Accordingly, the network between cell sites and mobile core is also divided into three segments: fronthaul, midhaul and backhaul. To address the strict bandwidth and latency requirements in 5G-RAN, 3GPP has proposed multiple functional spilt options, typically listed from option 1 to option 8, as shown in Fig. 2. In this paper we consider that the baseband processing entities (CU, DU and RU) are connected via two interfaces, F1 and Fx, as the split at Option 2 (interface F1) and Option 7 (interface Fx) have been selected by ITU [14] as the standard split options. According to [13], different baseband functions have various characteristics in terms of bandwidth requirement and processing complexity. The bandwidth, processing and latency requirement of CU, DU and RU can be summarized as follows.

Transport bandwidth: In the cell level, the F1 interface scales and dynamically varies with the air interface traffic load, so midhaul bandwidth scales with traffic load as backhaul and requires only slightly more bandwidth than backhaul. The fronthaul bandwidth at Fx also varies with the air interface traffic load but requires higher rates (typically by up to an order of magnitude compared to backhaul). In the user level, the bandwidth requirement of a user is related to the resource blocks (RB) occupied by this user in the carrier spectrum.

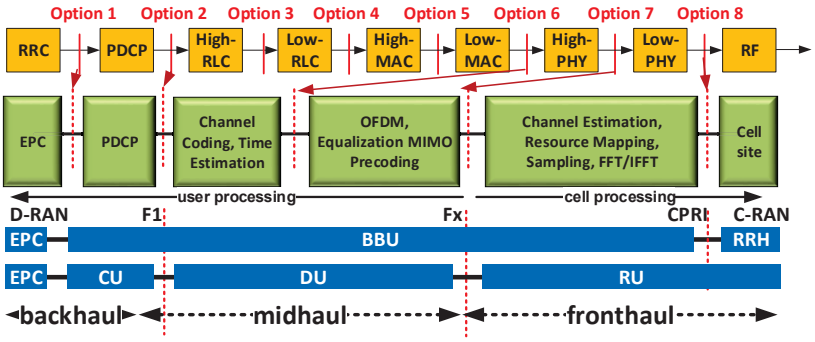


Fig. 2: Example of Function Split Options

Processing complexity: The processing complexity of baseband functions can be measured in Giga Operation Per Second (GOPS) [19]. The processing complexity of channel estimation, resource (de)mapping, FFT/IFFT is related to carrier bandwidth and number of antennas (called cell-processing functions), whereas the processing complexity of MAC/RLC/PDCP layer and the other functions in the PHY layer are related also to the traffic load besides carrier bandwidth and number of antennas (called user-processing functions). According to 3GPP [12], for option 2, RRC and PDCP are in the CU, while RLC, MAC, part of physical layer are in the DU. For the option 7 OFDM and MIMO precoding reside in the DU, FFT, resource mapping and RF resides in RU. So the processing complexities of DU and CU are all user load dependent. The details for the calculation of bandwidth requirement between CU, DU, RU and processing complexity of each entity used in this paper can be found in [13].

Transport latency: Transport latency requirements for backhaul and midhaul links are determined by service latency requirements, i.e. around 10 msec for eMBB, about 1 msec for URLLC and ranging from 1 msec to several 10 msec for mMTC [16]. For the fronthaul links at Option 7, the latencies are determined by the requirements of the RAN technology. To satisfy the HARQ processing latency requirement, a total round-trip latency budget of $\mathrm{RTT}_{BBU-RRH} = 3$ ms is available between a BBU and its corresponding RRH. It means that the NACK/ACK should be transported on the fronthaul link within hundreds of microseconds [15]. As we know the HARQ processing function is located in the low-MAC layer and MAC layer is in the DU, so latency requirement between DU and RU should be the same as HARQ processing latency requirement. According to [17], the reference values of the latency in different parts of networks are given as follows:

– Signal processing latency: The BBU completes the processing and send ACK/NACK usually within 2.75 msec. In the 3-layer architecture, no matter where the DU and CU are placed, the total processing can be seen as a fixed value.

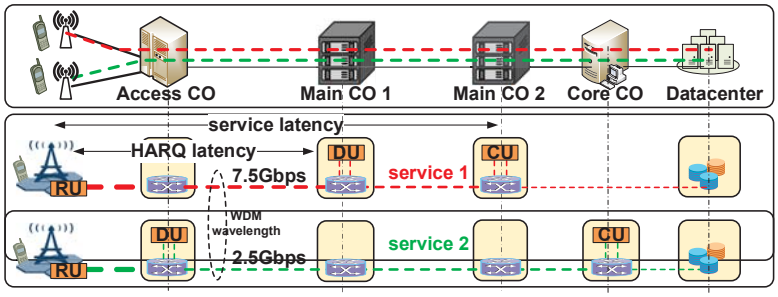


Fig. 3: Illustration of possible baseband function placement

– Signal propagation latency: For HARQ processing, propagation latency is related to the distance between RU and DU, and for the whole service, the propagation latency is related to the distance between the server and the RU (we ignore wireless transmission latency).
– Switching latency: We assume an active solution for the fronthaul, for OTN encapsulation, the latency is about 40 μsec, for the non-OTN encapsulation, the latency is about few μsec.
– Encapsulation (like CPRI) processing latency: Before the CU or DU transmit or receive the data, the data must be encapsulated or de-encapsulated with CPRI protocol. This procedure will cost around 10 μsec.

4 Joint DU/CU Placement in Metro Networks

4.1 Problem Statement

The DU/CU placement problem over metro/aggregation networks can be stated as follows. Given network topology, number of wavelengths per link and their line-rate capacity, set of traffic requests, maximum fronthaul latency (HARQ latency) and service latency, decide the placement of DUs and CUs in different COs that maximizes the consolidation of the baseband processing functions (i.e., minimizes the number of CO housing processing) under latency and capacity constraints.

The placement of CU/DU is not only restricted by the processing capacity and network capacity, but also subjected to the RAN latency and service latency constraints. Fig. 3 shows an example of CU/DU placement with two service requests from different cell sites. For service 1, the DU and CU are placed in main COs for a higher consolidation, whereas for service 2, restricted by the bandwidth capacity, the DU of service 2 must be placed in the access CO. Next, we will show the mathematic model of the DU/CU placement problem.

4.2 Model

1) ***Given sets and parameters***
V: set of COs
S: set of service requests
N: set of processing elements: RU, DU, CU, server
C_p^i: processing capacity of COs i(GOPS)
C_w: bandwidth of wavelength (Gbps)
G_n: computing resource requirement of baseband function n
B_n: bandwidth requirement between baseband function n and $n+1$
$T^{i,j}$: transport latency between CO i and j
T_{oe}: latency for switching and encapsulation
T_{HARQ}: latency requirement for HARQ processing
T_s: latency reqirement for service request
$K_{s,i}$: service s accessed into CO i
Max: a large positive value
2) ***Decision Variables***
$Y_i^{s,n}$: 1,if the function n of service s is located in CO i, otherwise 0
$X_{s,n}^{i,j,w}$: 1, if the function n and $n+1$ of service s is located in the CO i,j separately using wavelength w
D_i: 1, if the baseband processing functions (either CU or DU) are placed in CO i, otherwise 0
$H_{s,r}$: 1, if the service s is processed in the CO r
3) ***Objective***
Minimize the number of active COs:

$$\min\{\sum\nolimits_{i\in V} D_i\} \tag{1}$$

4) ***Constraints***
Routing:

$$Y_i^{s,n} = \begin{cases} K_{s,i}, & if\ n = 1 \\ 1, & if\ i = dest, n = N \end{cases}, \forall s \in S, i \in V, N \in N \tag{2}$$

$$\begin{aligned} &\sum\nolimits_{n\in N, i\in V, w\in W} X_{s,n}^{i,k,w} - \sum\nolimits_{n\in N, j\in V, w\in W} X_{s,n}^{k,j,w} \\ &\qquad = \begin{cases} -1, if\ K_{s,k} = 1 \\ 1, if\ k = 1 \\ 0,\ otherwise \end{cases} \quad \forall s \in S, k \in V, l \in L \end{aligned} \tag{3}$$

Baseband function placement:

$$\begin{aligned} Y_r^{s,k} \leqslant \sum\nolimits_{j\in V, w\in W, k\leqslant l\leqslant N-1} X_{s,n}^{i,j,w} \cdot M^{r,j} \leqslant Y_i^{s,k}, \\ \forall s \in S, k \in N, r \in V, K_{s,r} = 1 \end{aligned} \tag{4}$$

$$2Y_r^{s,k} \leqslant \sum\nolimits_{i\in V,w\in W,0\leqslant l\leqslant k-1} X_{s,l}^{i,r,w} \cdot M^{i,r} + \sum\nolimits_{j\in V,w\in W,k\leqslant h\leqslant N-1} X_{s,h}^{r,j,w} \cdot M^{r,j} \leqslant Y_r^{s,k} + 1, \quad \forall s \in S, k \in N, r \in V, K_{s,r} \neq 1 \tag{5}$$

$$D_i \leqslant \sum\nolimits_{s\in S,n\in N} Y_i^{s,n} \leqslant Max \cdot D_i, \ \forall i \in V \tag{6}$$

$$\sum\nolimits_{i\in V} Y_i^{s,n} = 1, \forall s \in S, n \in N \tag{7}$$

$$H_{s,r} \leqslant \sum\nolimits_{n\in N} Y_r^{s,n} \leqslant Max \cdot H_{s,r}, \forall s \in S, r \in V \tag{8}$$

Capacity:

$$\sum\nolimits_{s\in S,n\in N} G_n \cdot Y_i^{s,n} \leqslant C_p^i, \forall i \in V \tag{9}$$

$$\sum\nolimits_{s\in S,n\in N} B_n \cdot X_{s,n}^{i,j,w} \leqslant C_w, \forall i, j \in V, i \neq j, w \in W \tag{10}$$

Latency:

$$\sum\nolimits_{i,j\in V,w\in W} T^{i,j} \cdot X_{s,0}^{i,j,w} \leqslant T_{HARQ}, \forall s \in S \tag{11}$$

$$\sum\nolimits_{i,j\in V,n\in N,w\in W} X_{s,n}^{i,j,w} \cdot T^{i,j} + \sum\nolimits_{r\in V} H_{s,r} \cdot T_{oe} \leqslant T_s, \quad \forall s \in S \tag{12}$$

Additional bandwidth capacity constraint for overlay network:

$$\sum\nolimits_{s\in S,n\in N} X_{s,n}^{i,j,w} \leqslant 1, \forall i, j \in V, w \in W \tag{13}$$

Eqns. (2), (3) enable the routing of requests over the lightpaths, the source/ destination of a request is given and the baseband functions are flexibly placed in the intermediate nodes. Eqns. (4) and (5) restricts the services starting from the source node of a request and the links between the intermediate nodes. Eqn. (6) indicate that if the CO i is active. Eqn. (7) ensures that one baseband processing entity must be located in only one CO. Eqn. (8) indicates that if the service is processed in the CO r. Eqns. (9-11) is the capacity constraints of processing and bandwidth. Eqn. (12) ensures that the placement of DU is restricted by the HARQ processing latency. Eqn. (13) is for the service latency. Eqn. (14) is the additional bandwidth constraint for the overlay solution.

Table 1: No. Users of the Considered Geotypes

	Dense Urban	**Urban**	**Suburban**
Total Area of 32 sites $[km^2]$	8	22	160
Total Number of Users	2.4×10^4	2.2×10^4	8×10^4

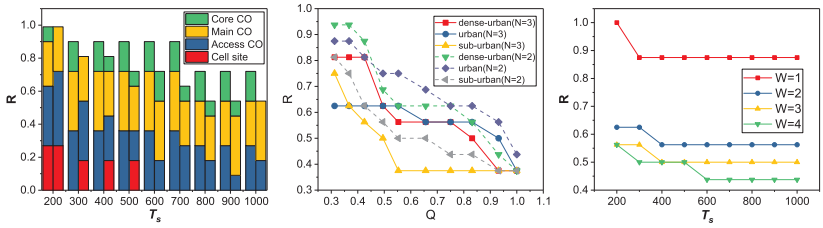
5 Illustrative Numerical Results

5.1 Evaluation Settings

We consider a 15-node WDM metro/aggregation network topology as shown in Fig. 4, under three different geographical type areas (geotypes) - Dense urban, Urban, and Sub-urban - with 32 cell sites distributed in that area. The total area of cell sites for different geotypes and the total number of users in the corresponding area are shown in Table 1 [18]. For the cell site, we assume a radio configuration with 20 MHz, 2x2 MIMO antenna, 64QAM modulation scheme and full system load, and the reference value for processing complexity [19] of different baseband entities and bandwidth requirements [13] of different network segments for the mentioned radio configuration is shown in Table 2. We assume that the resource blocks of a cell are uniformly allocated to all users, that are normally distributed in the whole area, so the processing complexity and bandwidth requirement is equally divided by all the users associated with the same cell site. For the transport network, the COs in different stages are equipped with different levels of baseband processing and switching capacity. COs are organized in a ring topology and are connected via bidirectional monofiber links, carrying W wavelengths at 10Gb/s. For the latency constraint, according to [17], maximum HARQ latency (RU-DU) is set to be 246 μs and maximum service latency is set to be 1000 μs. When below 246 μs, the HARQ latency requirement and service latency requirement are equal, and when above 246 μs, the HARQ latency is fixed at 246 μs. We define a performance metric R, called consolidation factor, which is used to evaluate multiplexing gain of function in our simulation. $R = N_{co}/N_{cs}$, N_{co} is the numbers of active COs and cell sites, N_{cs} is the number of cell sites. R=1 indicates no consolidation, that is all the baseband functions are located in the cell sites, whereas $1/N_{cs}$ indicates the highest consolidation degree.

Table 2: The value of processing complexity (GOPS) and bandwidth requirement (Gbps) with radio configuration of 20 MHz, 2×2 MIMO, 64QAM (Downlink)

Baseband entities	**Value**	**Network Segment**	**Value**
RU	48.1	fronthaul	0.97
DU	9.1	midhaul	0.299
RU	18.7	backhaul	0.299



(a) (b) (c)

Fig. 4: (a) R value for increasing T_s, and different RAN architectures (2-layer vs 3-layer, Q = 75%); (b) R values for increasing Q under different geotypes (overlay case; latency = 800 μs); (c) R values for the increasing W (OTN case; Q = 50%)

5.2 Evaluation results

Fig.4(a) shows the relationship between R and service latency requirement T_s, for the 2-layer and 3-layer RAN architecture (left and right bars, respectively), when considering overlay transport technology, urban geotype case and actual processing capacity set to be 75 % of maximum processing capacity. The value of R represented as the sum of four contributions, corresponding to consolidation degree in each stage of the network. With the increase of T_s, CUs and DUs can be consolidated in fewer COs to minimize the number of active COs, due to the fact that less stringent latency requirement allows the services to route along a longer path and end up in the COs of higher stages. For example, when the $T_s \geq$ 600 μs, there is no baseband function located in the cell sites (red color). We can also find that R value of 3-layer RAN architecture is less than the one of 2-layer RAN architecture, because flexible functional split divides the traditional BBU into multiple parts which can be deployed into COs more flexibly.

Fig.4(b) shows R as a function of $Q = \text{GOPS}_a/\text{GOPS}_m$ for the different geotypes and RAN architectures when considering $T_s = 800$ μs and overlay case. $\text{GOPS}_a/\text{GOPS}_m$ is the total actual processing capacity of all the COs divided by the maximum processing capacity in the networks. We can observe that higher consolidation is obtained when Q is increasing, especially for the suburban case. Because with the increasing processing capacity of COs, the DU and CU can be finally placed in the main or core COs when the latency requirement is relatively relaxed. For example, when Q goes near to 1, the value of R reaches the lowest level among almost all the scenarios. Note also that, due to the lower number of requests in the suburban area, the DUs/CUs are more easily deployed in the COs of higher stage. From this result, we can also find that 3-layer architecture benefits more in terms of consolidation that 2-layer architecture because of the flexibility in placement.

Fig.4(c) shows the relationship between R and network capacity in terms of number of wavelengths W. We can find that the relation between R and la-

tency depends on the bandwidth constraint. When the number of wavelengths is limited, no matter how loose the maximum latency requirement is, consolidation factor R will not decrease when it arrives at a critical point. This can be explained since when if enough wavelengths are provided, the provision of baseband functions can be more flexible; whereas if the bandwidth between different COs is limited, the location of baseband functions tends to be closer to the cell sites.

6 Conclusion

In this work, we have modeled an optimized DU/CU placement problem for C-RAN deployment over a WDM metro/aggregation network and formalized it into an ILP model. Compared to the original 2-layer RAN architecture, we proved that the 3-layer RAN architecture has higher consolidation of baseband functions thanks to the increased placement flexibility. We also observed that: 1) looser latency constraint leads to a high degree consolidation of baseband functions; 2) the processing capacity of COs also influences the consolidation of baseband functions; 3) adopting overlay transport solution can lead to a higher baseband function consolidation. In this work, the functional split options between RU, DU and CU are fixed. In the future, we will investigate the DU/CU placement problem when the functional split options is flexible, so the relation between flexible DU/CU placement and flexible functional split option needs to be jointly evaluated. Also, the heuristics for this problem will be proposed for the realistic scenario with larger network topology. Moreover, compared to the static DU/CU placement, dynamic DU/CU allocation problem according to real-time service requests from mobile users is worthy to be investigated.

Acknowledgment

This work was supported by the National Nature Science Foundation of China Projects (No. 61871051, 61771073), the Nature Science Foundation of Beijing project (No. 4192039), the fund of State Key Laboratory of Advanced Optical Communication Systems and Networks, China, No.2019GZKF5, and China Scholarship Council Foundation.

References

1. E. Hossain and M. Hasan, "5G cellular: key enabling technologies and research challenges," IEEE Instrumentation & Measurement Magazine, vol. 18, no. 3, pp. 1121, Jun. 2015.
2. 5G White Paper, Next generation Mobile Network (NGMN) Alliance, White paper, Feb. 2015. [Online]. Available: https://www.ngmn.org
3. A. Pizzinat, P. Chanclou, F. Saliou, and T. Diallo, "Things you should know about fronthaul," IEEE/OSA Journal of Lightwave Technology, vol. 33, no. 5, pp. 10771083, Mar. 2015.

4. J. Zhang, Y. Xiao, D. Song, L. Bai and Y. Ji, "Joint Wavelength, Antenna, and Radio Resource Block Allocation for Massive MIMO enabled Beamforming in a TWDM-PON based Fronthaul," IEEE/OSA Journal of Lightwave Technology. doi: 10.1109/JLT.2019.2894152.
5. X. Wang, S. Thota, M. Tornatore, H. Chung, H. Lee, S. Park, and B. Mukherjee, "Energy-Efficient Virtual Base Station Formation in Optical-Access-Enabled Cloud-RAN," in IEEE Journal on Selected Areas in Communications, vol. 34, no. 5, pp. 1130-1139, May 2016.
6. J. Zhang, Y. Ji, X. Xu, H. Li, Y. Zhao, and J. Zhang, "Energy Efficient Baseband Unit Aggregation in Cloud Radio and Optical Access Networks," IEEE/OSA J. Opt. Commun. Netw. 8, 893-901 (2016)
7. H. Yu, J. Zhang, Y. Ji and M. Tornatore, "Energy-efficient dynamic lightpath adjustment in a decomposed AWGR-based passive WDM fronthaul," IEEE/OSA Journal of Optical Communications and Networking, vol. 10, no. 9, pp. 749-759, Sept. 2018.
8. F. Musumeci, C. Bellanzon, N. Carapellese, M. Tornatore, A. Pattavina, and S. Gosselin, "Optimal BBU Placement for 5G C-RAN Deployment Over WDM Aggregation Networks," IEEE/OSA J. Lightwave Technol. 34, 1963-1970 (2016)
9. Y. Li, M. Bhopalwala, S. Das, J. Yu, W. Mo, M. Ruffini and D. C. Kilper, "Joint Optimization of BBU Pool Allocation and Selection for C-RAN Networks," 2018 Optical Fiber Communications Conference and Exposition (OFC), San Diego, CA, 2018, pp. 1-3.
10. J. Liu, S. Zhou, J. Gong, Z. Niu and S. Xu, "Graph-based framework for flexible baseband function splitting and placement in C-RAN," 2015 IEEE International Conference on Communications (ICC), London, 2015, pp. 1958-1963.
11. Y. Xiao, J. Zhang, Y. Ji, "Energy efficient Placement of Baseband Functions and Mobile Edge Computing in 5G Networks," Asia Communications and Photonics Conference (ACP), pp. 1-3, Oct 2018.
12. 3GPP TR 38.801 V14.0.0 (2017-03), Radio access architecture and interfaces (Release 14)
13. Small Cell Forum, "Functional splits and use cases for small cell virtualization," Jan. 2016.
14. G.sup.5GP, "5G Wireless Fronthaul Requirements in a PON Context," ITU-T (expected to be released by October 2018)
15. 3GPP TS-36.213 (Physical layer procedures). (2015). [Online]. Available: http://www.3gpp.org
16. Nokia, White Paper 5G new radio network, Document code SR1803023634EN (April)
17. Netmanias,"Fronthaul Size: Calculation of maximum distance between RRH (at cell site) and BBU (at CO)," Tech-Blog.
18. Deliverable 3.3: Analysis of Transport Network Architectures for Structural Convergence, COnvergence of fixed and Mobile BrOadband access/aggregation networks- COMBO Project, Tech. Rep., Jul. 2015. [Online]. Available: https://www.ict-combo.eu
19. B. Debaillie, C. Desset and F. Louagie, "A Flexible and Future-Proof Power Model for Cellular Base Stations," 2015 IEEE 81st Vehicular Technology Conference (VTC Spring), Glasgow, 2015, pp. 1-7.