Impact of Slice Granularity in Centralization Benefit of 5G Radio Access Network

Nabhasmita Sen
Department of Computer Science and Engineering
Indian Institute of Technology Hyderabad, India
cs17resch11001@iith.ac.in

Antony Franklin. A
Department of Computer Science and Engineering
Indian Institute of Technology Hyderabad, India
antony.franklin@iith.ac.in

***Abstract*—5G Mobile network will reap the benefits from key technologies like Software Defined Networking and Network Function Virtualization. Cloud Radio Access Network architecture (Cloud RAN) is proven to be a promising architecture, but fully centralized Cloud RAN imposes a great bandwidth requirement in the fronthaul link. Different functional split options for 5G RAN have been proposed which lead to a trade-off between centralization and bandwidth requirement. Functional split at different granularity such as per cell, per logical network (slice), per user, or per bearer, have been an area of interest. To explore the effect of slice granularity in adaptive splits for slices, we formulate slice centric functional split in 5G RAN as an ILP to maximize centralization of baseband processing. By varying the slice granularity from macro slicing to micro slicing, we observe how slice centric split can impact centralization benefit of the network. We show that with increasing slice granularity slice centric split can render more centralization benefit in some scenarios but a trade off exists between centralization benefit and migration cost in the network which should be carefully considered in real deployment scenario.**

***Index Terms*—Cloud RAN, Functional split, Centralized Unit (CU), Distributed Unit (DU), Network Slice.**

I. Introduction

5G radio access network is going to encompass miscellaneous technologies to cope up with the huge traffic demand and heterogeneous requirement of cellular network users. Along with providing various services to satisfy users, energy-efficient resource utilization is one of the main concerns from the perspective of operators. To increase network coverage with improved spectral efficiency, small cells came into picture but the benefits come at the cost of high capital expenditure (CAPEX) and operational expenditure (OPEX) as more base stations are deployed to serve ever increasing number of users. Even though load of the base stations vary with space and time i.e., spatio-temporal load variation [1] all the base stations must support full load in order to support peak traffic. To mitigate the effect on the operator's cost, Cloud RAN [2] proposes the execution of baseband functionalities of many base stations in a common central location. There are many benefits of this centralization of baseband functionalities [1] such as, higher utilization of processing resources, lower power consumption, easy sharing of signaling between base stations, cost saving on CAPEX and OPEX etc., and this is known as centralization benefit of cloud RAN. Though there are many benefits proposed in cloud RAN, as the size of the network grows, one of the main problems of fully centralized Cloud RAN is the cost of fronthaul link between Baseband Unit (BBU) and Remote Radio Head (RRH). Centralizing all functionalities of the baseband processing requires a high capacity fronthaul that incurs a huge deployment cost [2]. To alleviate this issue, different split points in the baseband function chain were introduced which are known as functional splits in RAN which can relax the fronthaul bandwidth requirement depending on the split point. Based on the split point, the centralization amount varies and finding an appropriate degree of centralization according to available resources has been a focus in 5G RAN research. The introduction of Software Defined Networking (SDN) and Network Function Virtualization (NFV) results in a rapid change in 5G mobile network architecture compared to the traditional network. With the help of NFV in 5G RAN, the BBU functions can be containerized and performed in the cloud while flexibly scaling resources for processing Cloud RAN functions. One of the most important aspect of 5G mobile network is Network slicing which is going to tremendously impact the way underlying resources are utilized. Slicing allows treating one physical network as multiple logical networks which can share resources in an efficient way. A varied range of services with a diverse set of requirements like data rate, connectivity, delay, etc., will be there in 5G. Based on the demand of these services they are classified into three main categories. Enhanced Mobile Broadband (eMBB) services which demand high data rate, massive Mobile type communication (mMTC) which requires connectivity for a large number of devices, and Ultra Reliable Low Latency Communication (URLLC) as the name suggests, needs low delay and highly reliable communication. Macro and micro slicing is a new area of consideration where granularity of slice has been one of the crucial aspects where macro and micro slicing refers to the coarse grained and fine grained slicing respectively. In course grained slicing there are few slices which comprise all the user flows in the network and as the granularity becomes finer, the number of slices are increased in the network to accommodate the users. Slice at cell, service, user, or bearer level have been considered in [3] which describes slices of different level of granularity. Applying functional split at different granularity of the network has been proposed as possible option. To analyze

978-1-7281-5684-2/20/$31.00 ©2020 IEEE

the benefits and feasibility of applying functional splits in various granularity of network such as per cell, per slice, per user, or per bearer,

- We formalize the slice centric functional split as an optimization problem to maximize centralization in a fronthaul constrained network.
- We compare slice centric split with baseline DU centric split and show that slice centric split can render more centralization benefit in many scenarios.
- Considering different granularity of slices from macro to micro slicing regime, we show that as slice granularity increases, more centralization in the network can be achieved but the migration cost for changing functional split also increases in the network.

II. Related Works and Motivation

Selecting appropriate functional split in Cloud RAN has been focus of research for few years. Authors of [4] have proposed an ILP to minimize energy and bandwidth consumption under given capacity of network and user load. In [5], a virtual network embedding algorithm is proposed and heuristics are given to address flexible functional split in a dense network. Authors of [6] propose a co-operation scheme to manage functional split in a fronthaul constrained network where the algorithm chooses functional split adaptively for all the Distributed Units (DU) under the same Centralized Unit (CU). In [7], the authors propose an agile RAN architecture for implementation of user centric functional split to minimize energy and bandwidth consumption. However, the above mentioned works don't analyze how slice granularity can affect while applying functional split in the network. Authors of [8] have considered slice centric functional splits but didn't consider physical layer splits which can give better centralization if resources are available. In [9], slice granularity and resource availability has been considered, but it does not talk about functional split and centralization benefits related to slices. In [10], an ILP has been proposed for optimal resource placement, however, the impact of slice granularity is not an objective here. When same functional split is applied to the midhaul connected to a DU without considering slices, we call this as DU centric functional split or DU centric split and when slices are considered and functional splits are decided per slice, we call it as slice centric functional split or slice centric split, further when per user functional split is decided it is called user centric functional split or user centric split. Deciding DU centric split is simple but can be less efficient in utilization of resources whereas, user centric split can enable utilization of resources more efficiently, but it can be challenging for practical implementation. So, to analyse the impact of slice granularity in functional split, in this paper, we design slice centric functional split as an ILP and explore how different granularity of slice can impact centralization benefit of Cloud RAN and the trade off that exists between centralization benefit and migration cost of baseband processing function while changing functional split.

III. System Model

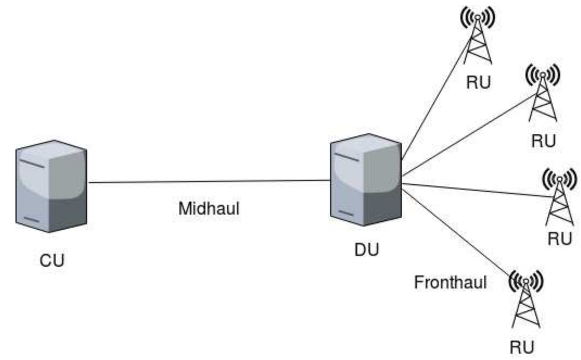


Fig. 1. RAN system model.

The three main components of Cloud RAN are Centralized Unit (CU), Distributed Unit (DU) and Radio Unit (RU) as shown in Fig. 1. CU is the central site where baseband processing of many base stations take place. DU also can have processing capability to perform baseband processing if needed and is kept at a remote site near to RU. RU is a lightweight unit whose function is to transmit and receive signals. In [3], different deployment options for these components are described. As depicted in Fig. 1, we consider that a CU is connected to a DU using a midhaul link where the DU supports many number of RUs with the help of short distance fronthaul links. In our model, the midhaul is shared by many services (such as broadband, IoT) [6] and the available bandwidth for mobile services varies with time [11]. So, on the basis of this variable midhaul bandwidth and network load, proper functional split should be selected. An SDN controller placed in CU keeps track of the available bandwidth, load, channel quality, etc., and assigns appropriate functional splits based on some objectives.

A. Baseband Processing functions

Baseband processing functions are the chain of functions performed by a base station in mobile network. These functions are of two different categories - Cell processing functions and User processing functions. Processing requirement of Cell processing functions depends upon cell parameters like bandwidth supported by the cell, number of antenna, etc., which are fixed for a particular cell. On the other hand, processing requirement of User processing functions varies with user load and depend upon the number of Physical Resource Blocks (PRB), channel quality, etc. In [12], 6 different functions are considered for RAN processing (Fig. 2). These functions are I. Higher layer functions, which comprises of Physical Data Convergence Protocol (PDCP), Radio Link Control (RLC) and Medium Access Control (MAC) Layers, II. FEC (Forward Error Correction), III. QAM (Quadrature Amplitude Modulation) and Antenna Mapping, IV. Resource Mapping, V. iFFT (inverse Fast Fourier Transform) and Cyclic prefix, and VI. P/S conversion and CPRI. As described in [7], processing functions II and III are considered as User processing functions and I, IV, V, and VI are considered as Cell processing functions.

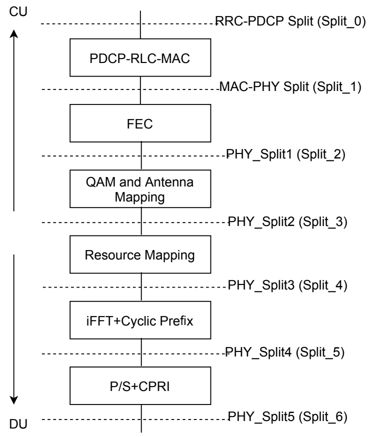


Fig. 2. Processing functions and different split points.

B. Functional Splits and Bandwidth Calculation

The different splits corresponding to these baseband processing functions are shown in Fig. 2. Split number from 0 to 6 are assigned from RRC-PDCP Split to PHY_Split_6 consecutively. Split_0 refers to scenario when all baseband functions are processed in DU that means no centralization and Split_6 refers when all functions are performed in the CU means full centralization. The seven different splits shown in Fig. 2 are of two different categories. Some are Cell splits and others are User splits [12]. Cell Splits have to be done for the whole cell together whereas User Splits can be done per user. When Cell split is done for a cell then that Cell split is enabled for all the users belonging to that cell, so all slices must have same split. But if in case User split is chosen, different users can have different functional split i.e., different slices can use different functional splits for them. Split_1, Split_2, and Split_3 are User Splits and others are Cell Splits. The bandwidth requirement for different functional splits, are calculated by using formulas given in [12].

C. Energy Calculation for Processing Functions and Centralization Benefit

For the approximation of consumed energy by a process, one of the most popular technique is to observe the number of CPU cycles consumed to perform that process and converting it to energy by knowing the system capacity. To measure the centralization benefit, we consider the amount of baseband processing that can be done in the CU as it can give a rough idea about energy consumption to perform those processes. Details of how the digital baseband processing functions consume energy and upon which factors they depend are given in [13]. Based on that, [7] has a model which we refer for energy calculation of different baseband processing functions as given in Table I. The A, B, L, and M are number of antenna, bandwidth, traffic load, and modulation index, respectively of the network and A_{ref}, B_{ref}, L_{ref}, and M_{ref} are number of antenna, bandwidth, traffic load, and modulation index, respectively in the reference scenario. The g_is are the processing requirement in Gigabit Operations Per Second (GOPS) which gives number of CPU cycles needed for different baseband functions in the reference scenario.

TABLE I
PROCESSING REQUIREMENT FOR EACH PROCESSING FUNCTION

Processing Function	Calculation
I	$g_1 * \frac{A}{A_{ref}}$
II	$g_2 * \frac{B}{B_{ref}} * \frac{M}{M_{ref}} * \frac{A}{A_{ref}} * \frac{L_i}{L_{ref}}$
III	$g_3 * \frac{B}{B_{ref}} * (\frac{A}{A_{ref}})^2 * \frac{L_i}{L_{ref}}$
IV	$g_4 * \frac{B}{B_{ref}} * \frac{A}{A_{ref}} * \Sigma_{i=1}^{N} \frac{L_i}{L_{ref}}$, N=Total users
V	$g_5 * \frac{B}{B_{ref}} * \frac{A}{A_{ref}}$
VI	$g_6 * \frac{B}{B_{ref}} * \frac{A}{A_{ref}}$

D. DU Centric Functional Split Vs. Slice Centric Functional Split

In 5G, the physical network can be considered as one or multiple logical network or slice. Functional split in traditional network means to have single functional split for the whole network connected to a midhaul and single DU (Fig. 1), which we call DU centric split. But, in a Slice centric split, according to the operator's requirement, functional split decisions can be taken per slice. To visualize slice centric functional split, we refer to an example in Fig. 3 which shows that the midhaul bandwidth is shared among three different slices present in the network and the slices can select different functional splits for them according to the available bandwidth in the midhaul.

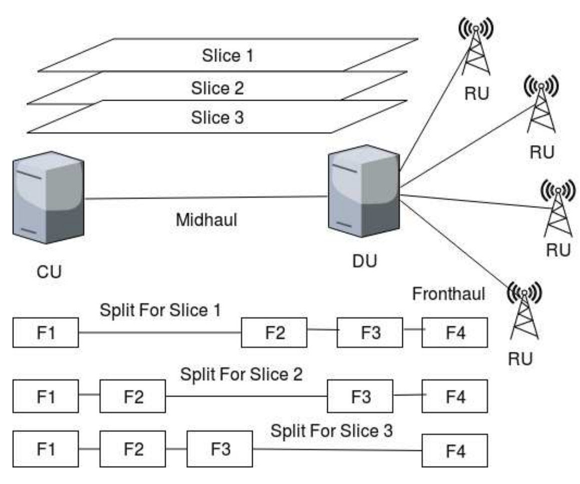
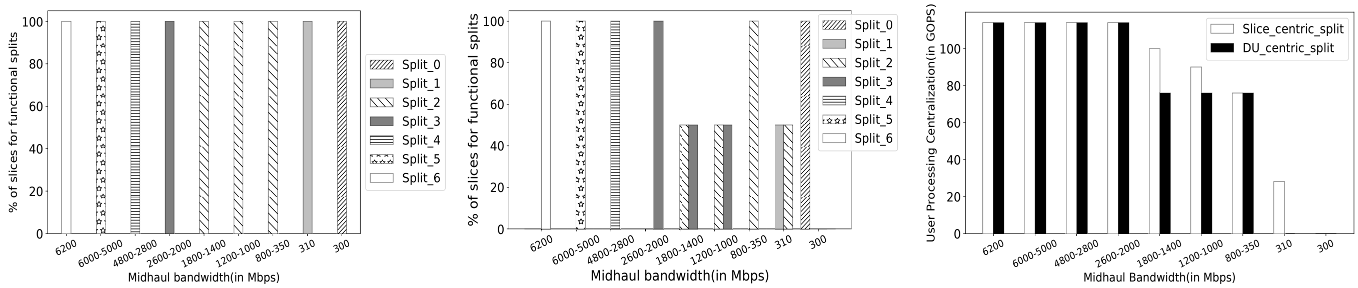


Fig. 3. Illustration of slice centric functional split.

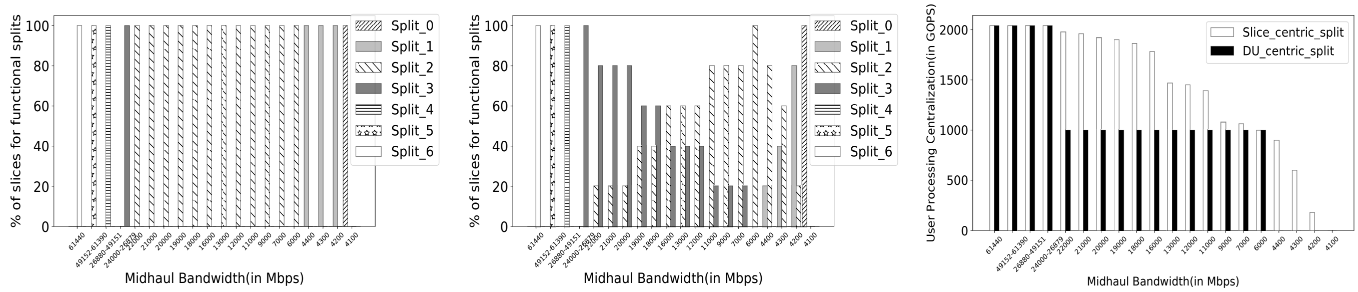
E. Objective and ILP Formulation

Centralization of baseband processing leads to better utilization of RAN resources. Centralization benefit in the network is referred to the amount of baseband processing that is performed in the CU. To analyze the benefit of slice centric split,



(a) DU centric split (b) Slice centric split (c) User processing centralization

Fig. 4. Comparison of DU centric and slice centric functional split (5 RUs and 2 slices).



(a) DU centric split (b) Slice centric split (c) User processing centralization

Fig. 5. Comparison of DU centric and slice centric functional split (50 RUs and 5 slices).

the objective of our ILP is to centralize baseband processing for all slices as much as possible by using different splits for different slices if necessary. With varying availability of midhaul bandwidth and load variation of slices, optimization problem selects functional splits for slices in such a way that the centralization of processing functions in CU can be maximized. The binary decision variables x_{ij} denotes functional split for different slices, x_{ij} is 1 if for slice i functional split j is activated otherwise it is 0 and z_k indicates Cell split which is 1 when k^{th} Cell split has been activated and 0 otherwise. The objective is,

$$Maximize : \sum_{i \in S} \sum_{j \in FS} x_{ij} CU_{ij} \tag{1}$$

Constraints for the optimization are,

$$\sum_{j \in FS} x_{ij} = 1, \forall i \in S \tag{2}$$

$$\sum_{k \in CS} z_k <= 1 \tag{3}$$

$$\sum_{i \in S} x_{ik} = |S| z_k, \forall k \in CS \tag{4}$$

$$\sum_{i \in S} \sum_{j \in FS} x_{ij} B_{ij} <= BW \tag{5}$$

$$x_{ij} \in \{0,1\}, \forall i \in S, \forall j \in FS \tag{6}$$

$$z_k \in \{0,1\}, \forall k \in CS \tag{7}$$

In Eqn. (1), CU_{ij} denotes the pre-calculated processing requirement in CU for slice i and functional split j. S, FS, and CS are the set of network slices, available functional splits, and the Cell splits, respectively. Eqn. (2) ensures one slice can have only one functional split. Eqn. (3) ensures at most one Cell split is activated for the network. Eqn. (4) denotes if Cell split is chosen then all slices must have the same Cell split. Eqn. (5) is the constraint that the bandwidth generated by all the slices should be less than or equal to available midhaul bandwidth, where B_{ij} is the pre-calculated bandwidth requirement in midhaul for slice i and functional split j and BW is the available bandwidth in the midhaul link at the time of optimization. Eqn. (6) and (7) ensures that x_{ij} and z_k can only take binary values. We assume that the midhaul capacity can support the lowest functional split i.e., RRC-PDCP split and all split support service requirement of slices. Also, the CU and DU have the processing capacity to support all the functional splits for the network.

IV. Simulation and Results

We take the baseline as DU centric functional split where one split is chosen for the midhaul between a DU and CU i.e., same functional split is selected for the whole network connected to a DU. We observe how a slice centric functional split can render more centralization benefit than DU centric split. Later, we check the impact of slice granularity in centralization benefit of the network. Optimization results are collected with the help of Gurobi optimizer.

A. Slice Centric split vs. DU Centric split

To find the centralization benefits of slice centric functional split over traditional functional split, we consider one CU is connected to one DU with the help of midhaul link which in turn connected to five RUs. All RUs are considered as SISO supporting 20 MHz channel bandwidth. Two types of slices are considered in the network, $Slice1$ with light traffic (such as mMTC) and $Slice2$ which requires heavy traffic requirement (such as eMBB). Users of $Slice1$ and $Slice2$ are given 4 and 16 PRBs, respectively and for all users, highest MCS is assumed. The total traffic of a slice consists of the traffic of all the users belonging to that slice. Each RU is connected with 10 users and out of which 70% of the users belong to $Slice1$ and 30% of the users belong to $Slice2$. Split_1, Split_2, and Split_3 are User splits and other splits are Cell splits (Fig. 2). We recall that when Cell split is activated, all the slices will have the same split but when User split is activated slices can have different User splits for them.

I. Available bandwidth in the midhaul link vs percentage of slices for different functional splits

We refer to Fig. 4 to analyze functional split assigned to slices according to midhaul bandwidth. In Fig. 4a and Fig. 4b, x-axis denotes different range of available bandwidth in the midhaul and y-axis denotes the percentage of slices for each functional split. As the bandwidth availability decreases, next highest possible split is chosen for the slices. For the Cell splits, users of both the slices have to use the same split (Fig. 4), as Split_6, Split_5, and split_4 are Cell splits, both the slices are having the same split. But, as soon as the bandwidth decreases further, different User splits are assigned to different slices. So, we can see that for some ranges of bandwidth, different slices are getting different splits (Fig. 4b). If DU centric split was done, both the slices would have got the same split (as in Fig. 4a) even though bandwidth could support different split for the slices. When available bandwidth becomes very low such that it can only support the lowest split then Split_0 is assigned to both the slices.

II. Centralization benefit in slice specific split

Here, we show the difference in processing centralization between DU centric split and slice centric split. As Cell processing functions are done for all slices together, only User processing functions can create difference between DU centric and slice centric split. So, Fig. 4c compares User processing function centralization in Slice centric split with DU centric split. Reference processing requirement is calculated with the energy model discussed in Section III. From Fig. 4, we observe that when same split is chosen for all the slices then same centralization is achieved in slice centric and DU centric split, But, in other cases when slices are able to get different splits for them, slice centric split can achieve more centralization than DU-centric split.

III. For a large site with more number of RUs and slices

In this case, we consider a large network scenario to observe centralization benefits between DU-centric split and slice centric split varying network size and number of slices. In this case, one CU is connected to one DU and 50 RUs are connected to that DU and each of the RUs having 10-12 number of users connected to it. There are 5 slices which comprises of all the users in the network. Fig. 5a and Fig. 5b show the percentage of splits in the network in case of DU and slice centric split and Fig. 5c shows User processing centralization benefit of slice centric split over DU centric split. From Fig. 5c, it can be observed that for a large range of midhaul bandwidth, Slice centric functional split renders better centralization benefit than DU centric split.

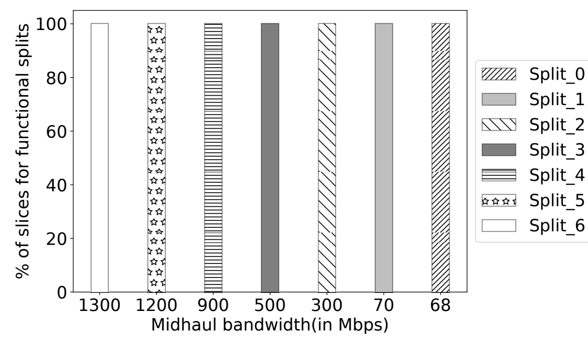


Fig. 6. Functional split changes in DU centric split (1 slice).

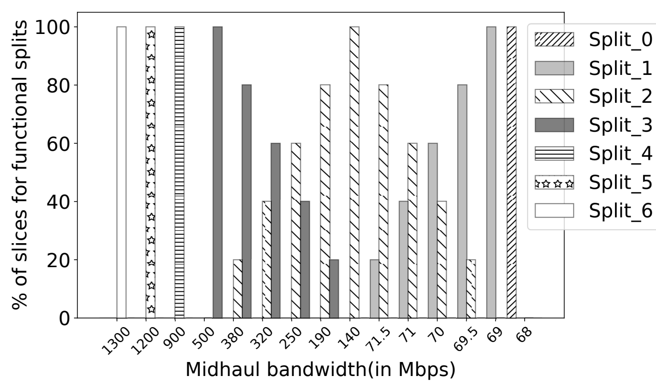


Fig. 7. Functional split changes in slice centric split (5 Slices).

B. Impact of slice granularity in Slice centric functional split

In this section, it is observed how increasing the granularity of slices can affect the centralization in the network. Even in a small network the difference can be visible. So, we consider the scenario where CU is connected to one DU which has one RU connected to it and there are 10 users connected to that RU. We assign the users 8 PRBs each per TTI and vary the number of slices. We increase the granularity of slices from all users in one slice (coarse grained) to one user per slice (fine grained). Functional split for all users in one slice will be same as DU centric split, whereas split for one user per slice is user centric split. Fig. 6, Fig. 7, and Fig. 8 show how different functional splits are chosen by slices when available bandwidth in the midhaul varies. As we go from DU-centric split (Fig. 6) and slice centric split with 5 slices (Fig. 7) to

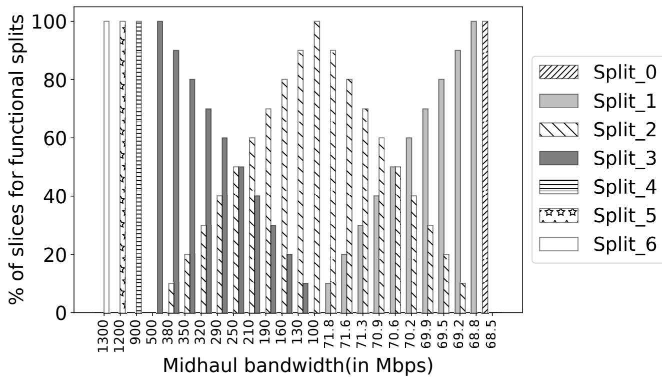


Fig. 8. Functional split changes in user centric split (10 Slices).

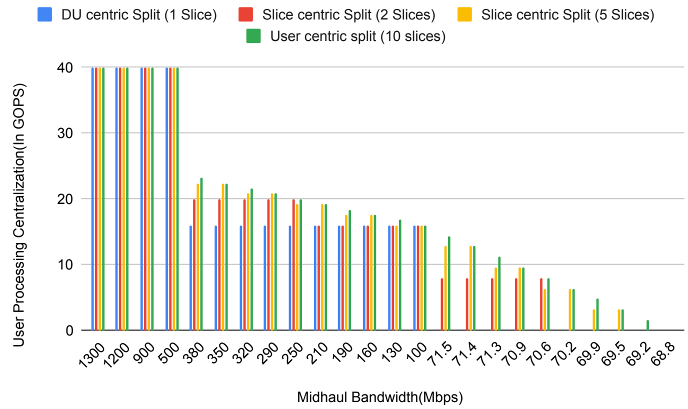


Fig. 9. Difference in processing centralization for different slice granularity.

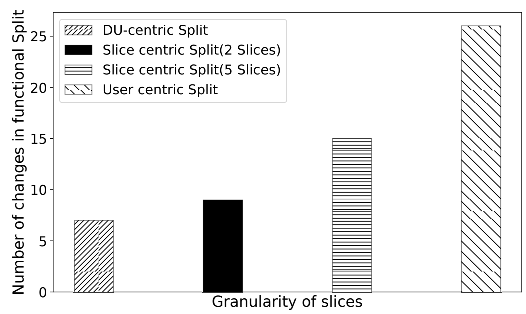


Fig. 10. Increasing slice granularity vs. frequency of functional split change.

user-centric split (Fig. 8), we see the number of times change in functional split occurs increases in the same range (1300-68 Mbps) of midhaul bandwidth. We compare DU centric and user-centric split with slice centric split (with different number of slices) and compare their processing centralization. As the Cell splits render the same centralization in all the cases and only User splits brings all the differences, we compare centralization achieved from User processing functions. In Fig. 9, it is observed that an increasing number of slices can give more processing centralization benefits in most cases for given midhaul bandwidth but this is not monotonous i.e., only increasing granularity does not ensure more centralization. For example, in Fig. 9, when midhaul bandwidth is 250 Mbps, it is observed that in slice centric split with 5 slices the processing centralization is less than slice centric split with 2 slices. The explanation for this as follows. In case of 2 slices, all the users (10 users) are divided into 2 slices where each slice is having 5 users, on the other hand in case of 5 slices each slice is having 2 users. When midhaul bandwidth is 250 Mbps, it can support traffic from maximum of 5 users for the highest possible split (each user having 8 PRBs). So in case of 2 slices, it supports 1 slice with 5 users for the highest possible split, but in case of 5 slices only 2 slices can be supported. Because, if it considers another slice for the highest possible split, total traffic will come from 6 users which the midhaul cannot support. So, it supports 2 of those slices i.e., supporting 4 users for a higher split and it gives the idea that not only granularity but the size of the slice also matters in case of attaining centralization in slice centric split.

As the midhaul bandwidth or network load varies ([11]) the functional splits changes in the network. Even if we fix the load and vary only the midhaul, from Fig. 10, we see how slice granularity can impact frequency of split change in the network. We plot the number of times functional split is changing for different number of slices in the same bandwidth range. It can be noticed that increasing number of slices can increase the frequency of change in functional splits in the network as even lesser difference in available bandwidth can cause a change in the functional split. Fig. 10 shows how migration cost (expressed as the number of times functional split changes) due to change in the functional splits will increase in case of user-centric functional split as the slice granularity increases from DU-centric split. From Fig. 9 and Fig. 10, it can be seen that though more centralization benefits can be achieved by increasing the number of slices, as the granularity of slice increases the frequency of change in functional split also increases. Increase in change in functional split may lead to service disruption or additional overhead in managing the network. Hence, there is a trade-off between centralization benefit and function migration cost while considering different slice granularity and this should be carefully considered in case of real deployment scenario.

V. Conclusion

In this paper, we propose slice centric functional split for varying available midhaul bandwidth as an optimization problem with the objective of maximizing the centralization of baseband processing. We select appropriate functional split for each slice so that the centralization of processing functions in the network can be maximized. By taking different granularity of slices, we show that adaptive selection of functional split for different slices can render more processing centralization than selecting the same functional split for all the slices in the network. We also show that increasing granularity of slices and even size of the slices can impact centralization benefit and processing function migration cost while implementing slice centric functional split. So, slice granularity should be carefully considered while selecting splits dynamically.

REFERENCES

[1] China Mobile Research Institute, “White paper: C-ran the road towards green ran”, Tech. Rep., 2011.

[2] A. Checko et al., “Cloud ran for mobile networks—a technology overview”, in *IEEE Communications Surveys Tutorials*, 2015.

[3] NGMN, “Ngmn overview on 5g ran functional decomposition”, Tech. Rep., 2018.

[4] X. Wang, A. Alabbasi, and C. Cavdar, “Interplay of energy and bandwidth consumption in cran with optimal function split”, in *2017 IEEE International Conference on Communications (ICC)*, 2017.

[5] D. Harutyunyan and R. Riggio, “Flex5g: Flexible functional split in 5g networks”, in *IEEE Transactions on Network and Service Management*, 2018.

[6] A. Marotta, D. Cassioli, K. Kondepu, C. Antonelli, and L. Valcarenghi, “Efficient management of flexible functional split through software defined 5g converged access”, in *2018 IEEE International Conference on Communications (ICC)*, 2018.

[7] S. Matoussi, I. Fajjari, S. Costanzo, N. Aitsaadi, and R. Langar, “A user centric virtual network function orchestration for agile 5g cloud-ran”, in *2018 IEEE International Conference on Communications (ICC)*, 2018.

[8] B. Ojaghi, F. Adelantado, E. Kartsakli, A. Antonopoulos, and C. Verikoukis, “Sliced-ran: Joint slicing and functional split in future 5g radio access networks”, in *IEEE International Conference on Communications (ICC)*, 2019.

[9] U. C. Kozat and A. C. K. Soong, “On the impact of slicing granularity on the availability and scalability of 5g networks”, in *IEEE International Conference on Communications (ICC)*, 2019.

[10] A. De Domenico, Y. Liu, and W. Yu, “Optimal computational resource allocation and network slicing deployment in 5g hybrid c-ran”, in *IEEE International Conference on Communications (ICC)*, 2019.

[11] A. Marotta, D. Cassioli, K. Kondepu, C. Antonelli, and L. Valcarenghi, “Exploiting flexible functional split in converged software defined access networks”, in *IEEE/OSA Journal of Optical Communications and Networking*, 2019.

[12] Small Cell Forum, “Small cell virtualization functional splits and use cases”, Tech. Rep., 2016.

[13] C. Desset et al., “Flexible power modeling of lte base stations”, in *2012 IEEE Wireless Communications and Networking Conference (WCNC)*, 2012.