Selection and Dimensioning of slice-based RAN Controller for adaptive Radio Resource Management

Emmanouil Pateromichelakis, Chenghui Peng

Huawei Technologies, European Research Center, Munich Germany

***Abstract*—This paper discusses the adaptive cluster formation and controller selection from a set of access nodes, given the RAN characteristics, the resource situation and the per slice QoS requirements, Subsequently, the adaptive placement of Radio Resource Management (RRM) functionalities to the RAN nodes and the interactions among functionalities are determined on per slice basis. By taking into account the slice requirements, the backhaul/access channel conditions and the traffic load, a central management entity assigns RRM functionalities to the controllers with different levels of centralization in order to meet the per Slice KPIs (throughput, reliability, latency). The controller, cluster and RRM split configuration problems are formulated and interpreted as three dependent graph-based sub-problems, where low complexity heuristic approaches, requiring low signalling, are proposed in this framework.**

***Index Terms*—Network Slices, Radio Resource Management**

I. INTRODUCTION

The explosive growth in capacity and coverage demands emerged the evolution of traditional Radio Access Networks (RANs) towards highly densified and heterogeneous deployments as foreseen in some 5G scenarios [1]. An exemplary use case, as defined in [2], is the dense urban scenario which assumes macro-cells (or macro-Transmit and Receive Point (TRP) as defined in 3GPP) and mixture of planned and un-planned small cells (or Micro- Transmit and Receive Points (TRPs) as defined in 3GPP) under the macro-cell umbrella, where both ideal and non-ideal backhaul can be present. In such environments, centralized solutions are highly required for optimal performance. This can be feasible solution for C-RAN deployments; however such feature is not present in D-RAN (see 5G RAN deployments 1 and 2 in Fig. 1). In 5G RAN, use cases originating from vertical industries (e.g. automotive, e-health, smart grid etc.) will be considered as drivers for 5G requirements. 5G network should be able to adapt to their needs in terms of latency, reliability, security, QoS, etc. To this end, the introduction of network slices, which are logical end-to-end sub-networks corresponding to different verticals, is envisioned as a key 5G feature [1]. Network slices might impact the RAN design and RRM is one of the key aspects which will be affected. Different slices aim at different goals e.g. throughput, or latency or reliability. This affects how RRM functions work and also where these functions can be placed

In dense heterogeneous RAN, multiple limitations for backhaul / access might require certain handling of resource management. In particular, non-ideal wireless backhaul between RAN nodes can be a limiting factor and will require extra RRM for the backhaul part. To this end, joint backhaul / access optimization can be used to meet high throughput requirements for throughput demanding services. Another important factor is the excessive signalling which will be required in Dense Urban heterogeneous scenarios (with macro, numerous overlapping small cells) for wireless backhaul (BH) & access measurements. This is going to be more crucial by new RRM functional interactions which will be added by performing fast scheduling decisions in small cells (could be the case in Ultra Reliable and Low Latency Communications (URLLC).

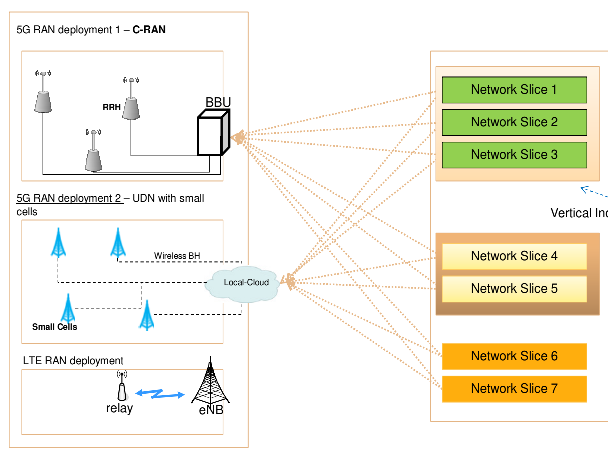

Figure 1 Slice-aware RAN deployments

The paper discusses a new concept which introduces the cluster formation from a set of access nodes, given the RAN characteristics (user/cell density, average mobility, and backhaul conditions) and the slice requirement. Subsequently, this involves also the selection, from a set of access points, a node to act as a slice-aware RRM controller in order to perform RRM functionalities with different levels of centralization.

The remainder of this paper is organized as follows. Section II presents the state-of-the-art on slice-aware RRM and categorization based on placement on RRM functions. Section III discusses the system model and problem description. Moreover, Section IV describes the solutions framework, which includes the mechanism for controller selection, cluster formation and RRM split configuration. Finally, Section V presents the signalling overhead and evaluation results and Section VI concludes our findings.

II. Radio Resource Management Considerations For 5G RAN

A. *Slice-aware RRM*

In Slice-aware RAN, different RRM functionalities and placements are required to ensure that service-tailored KPIs are met. 5G tight KPI requirements (esp. for URLLC and connected cars) rely on fast and sophisticated RRM. Furthermore, different RRM procedures can be required for diverse Slices and in different time scales. In RAN, the slice requirements are shaped by the target KPIs and key RAN characteristics (like user mobility and user / cell densities. Mobility may affect the backhaul resource allocation, handover and interference management. Also, the user density can affect delays, signalling required, interference management and resource availability. For each slice, these effects shall require different actions from access node point of view, to meet the end-to-end KPIs. In other words, different resource management and control placement of functions might be required from slice to slice. An example can be given for the URLLC type of Slice (e.g. for vehicular safety) where distributed RRM is preferred (most RRM locally), so the threshold to centralize RRM will be high (e.g. For Centralized Mobility Control in case of high mobility). In eMBB, on the other hand, we need high Centralization. If non-ideal BH exists, we choose the RRM Split as centralized as possible, except if the mobility is very low and there is no overlapping between small cells.

As stated in state-of-the-art literature [1][3], network slices will allow for flexible functional placements and tailored network functions to meet the per slice SLAs. Hence, slice specific resource management [4] and isolation among slices, utilizing the same RAN is an open topic which is currently investigated. In this context, Inter-slice RRM can be defined as another functional block which dictates the RAN sharing and level of isolation / prioritization among Network Slices. However, the impact of slicing on RRM functions which can trigger their adaptive placement in RAN nodes in a semi-distributed manner is not yet discussed.

B. *RRM Classification and Splits*

Based on 3GPP LTE RRM functionalities and structure [5], given the RAN limitations, RRM can be grouped in 3 main groups given their output, their in-between interactions and the time scale they operate: 1) Slow RRM, which can trigger cell selection / re-selection, 2) Fast RRM, which can change the resource utilization / restrictions, 3) Basic RRM for bearer admission and control. For the D-RAN scenario [3], we can also have a fourth type which is about the wireless backhaul resource management and wireless topology handling. Moreover, given the level of centralization of RRM, we can observe three different RRM types. In particular:

1) ***Centralized RRM:*** RRM functions operate together in an entity for multiple access nodes in a group. This will provide fast and simple interaction between RRM functions, but on the other hand in HetNets, we need ideal backhaul for some fast RRM functions (e.g. CoMP, DRA). Moreover, signalling overhead can be very high in ultra dense environments. Furthermore for 5G systems [6], some solutions have been proposed that require controller for clusters of Hetnets (Cloud-based Resource pooling and management (C-RAN) [7]. There, resource pooling and centralized management of resources can provide high gain in terms of capacity. Nevertheless, this requires ideal backhaul / fronthaul and can be seen as challenging task for dynamic resource allocation in fast changing environments assuming also interference from other C-RAN Clusters.

2) ***Distributed RRM:*** In 3GPP LTE / LTE-A [5], we observe that RRM functions reside at the eNodeB. The main RRM functions are about Dynamic Resource Allocation (DRA), Interference Coordination (ICIC) and Connection Mobility Control (CMC), Radio Bearer Admission and Control (RAC, RBC), Energy Efficiency and Load Balancing (LB). In LTE RRM structure, there can be interactions between RRM functions. An example is the Cell on/off function, which might require input from the resource restrictions due to interference management, and its output will require handovers which will affect mainly CMC and LB. Since, all these functions reside at the eNodeB, there is no additional signalling specified in 3GPP for the RRM interactions.

3) ***Semi-Centralized RRM:*** Other studies [8][9], discuss two levels of RRM, denoted as Global and Local Scheduling. These studies mainly focus on centralized Interference Management and Load Balancing and distributed fast RRM functionalities. One of the main challenges in this case is the new interactions that will require additional signalling and complexity in various RAN nodes. Fig. 2 illustrates these interactions in case of heterogeneous RAN with small cells under a macro-cell umbrella. Here we observe that if RRM functionalities will be semi-distributed between the macro and the small cells, extra signalling will be required since the duplication of some functions in different nodes will require exchange of information between the involved nodes.

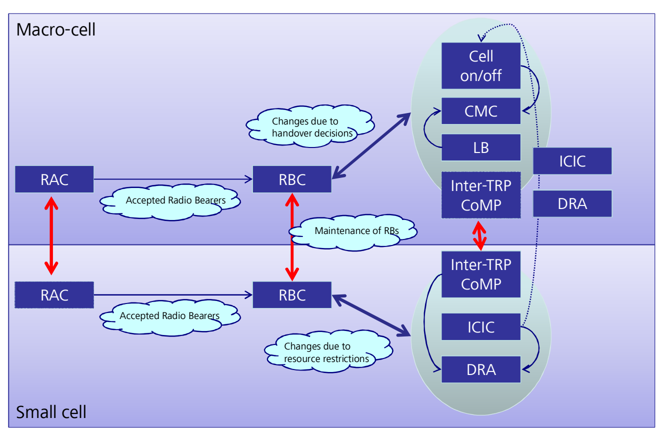

Figure 2 RRM possible interactions in 3GPP LTE-A with Dual Connectivity

Due to the tight interactions of RRM functions, a practical way to split the RRM is between Slow and Fast RRM given their time scale they operate. Hence, in this paper we discuss the split of RRM based on the categorization above. Below, in Fig. 3, different Splits can be illustrated. In our categorization, together with Slow and Fast RRM we can also have at the Controller new RRM functions for the wireless BH (topology RRM) and RRM for different Slices (Inter-Slice RRM).

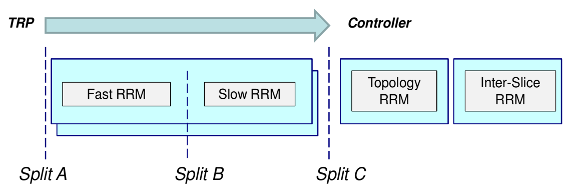

Figure 3 RRM Split and Grouping

Handling multiple and different resources in a dense urban 5G RAN, with different slice KPIs will require a controller to orchestrate the resource management and control between slices. Centralized Solutions will be required to meet required performance goals. In this context, a RRM Controller can be defined as a logical entity which abstracts a set of access network functionalities and coordinates a group of access nodes to facilitate the resource management and control. The benefit of such controller is that RRM can be optimized per KPI e.g. for throughput (using sophisticated Interference Management), mobility and reliability. However, the physical placement and dimensioning of RRM Controller plays important role for the efficiency of RRM. The high number of controllers will provide more granular RRM, however this might impose high delays for the communication between RRM Controllers due to the functional dependencies in case that these controllers reside in different entities. On the other hand, one flat RRM Controller for RAN will not allow for efficient slice-tailored RRM, since the centralization impact would be different from to slice to slice. Also, the backhaul capabilities will provide some limitations regarding the placement of such controllers.

So the problem can be formulated as the way to initially create clusters from a pool of access nodes or BSs (denoted as Transmit-Receive-Points (TRPs), then select and configure one or more nodes as their controller (s) (denoted as RRM Controller) and finally control the rest of the nodes of the cluster as the controlled entities (defined as Slave-TRPs). Subsequently, how to decide which RRM functions will be decided at the RRM Controller and which will be distributed at the other access nodes.

III. SYSTEM MODEL AND PROBLEM DESCRIPTION

The entire network encloses *l=1,2,..,L* TRPs which are connected to Network Management System (NMS). Let *s=1,2,..,S* slices in a RAN and NF_{total} the set the set of RRM network functions (NFs) which reside at RAN. Some of NFs can be slice-based and some can be common between all slices (NF_{common}), such that:

$$NF_{total} = (\cup_{s=1}^{S} NF_s) \cup NF_{common} \qquad (1)$$

In this study, we define four groups of RRM NFs, namely fast RRM (fRRM), slow RRM (sRRM), topology RRM (tRRM) and inter-slice RRM (isRRM). The criteria of selecting a NF to be in one of the two first groups are based on the periodicity t_{NF} and the slice requirements. For *tRRM* and *isRRM*, the grouping is pre-defined from the common NF set, based on the physical deployment and the slice awareness.

Definition 1: A $NF_s, \forall s \in S$ is considered to belong to fRRM, only if its periodicity t_{NF_s} is more granular than a pre-defined slice defined threshold $thresh1_s, \forall s \in S$

$$fRRM_s \coloneqq \{set\ of\ NF_s\colon\ t_{NF_s} \leq thresh1_s, \forall s \in S\}$$

Definition 2: A $NF_s, \forall s \in S$ is considered to belong to sRRM, only if its periodicity t_{NF_s} is less granular than a pre-defined slice defined threshold $thresh1_s, \forall s \in S$ and more frequent than a second slice-based threshold $thresh2_s, \forall s \in S$

$$sRRM_s \coloneqq \{set\ of\ NF_s\colon\ thresh1_s \leq t_{NF_s} \leq thresh2_s, \forall s \in S\}$$

Based on this grouping, three splits can be defined in this paper. *Split A* can be seen the case when Centralized RRM is performed ($fRRM_s of\ node\ i \in j, \forall i \neq j \in L$). *Split B* can be seen as the case of *sRRM* happening at the Central Unit (CU) and *fRRM* at the Distributed Unit (DU) ($sRRM_s of\ node\ i \in j, \forall i \neq j \in L$) .Also, *Split C* applies for the complete Distributed case ($sRRM_s of\ node\ i \in i, \forall i \in L$).

Let *G (V,E)* be the graph consisting of a set of *V* nodes (TRPs) and a set of *E* edges, such that $|V| = M_T$ and E edges that show the coordination potentials between TRPs. The graph is a weighted graph connecting each pair of TRPs *u, v* in the system through weighted edges $E(u, v), \forall u, v \in V$. The weight corresponds to the potential benefit of such coordination and can be seen as a function of the distance $d(u, v), \forall u, v \in V$, the backhaul channel conditions $BH(u, v), \forall u, v \in V$, the traffic load at the edges between these nodes $Load_{u,v}$ and a normalized weighting factor $Cs_{u,v}(s), , \forall u, v \in V, \forall s \in S$, which reflects the slice preference towards the utilization of this edge, assuming a set of *S* slices in the network. The weight of the edge can be interpreted as the function of these factors as:

$$W_E = \{w_{u,v} = \mathrm{F}\{d(u, v), BH(u, v), Load_{u,v}, \sum_{s \in S} Cs_{u,v}(s)\}, \quad \forall u, v \in V, s \in S\} \qquad (2)$$

This problem, as mentioned in previous section, can be seen a very complex combinatorial optimization problem with multiple sub-problems. To address this issue, in this work, we introduce an alternative graph theoretic formulation and we de-couple this into 3 sub-problems.

A. Cluster Formation Sub-Problem

Firstly the clustering can be translated to finding the maximal weighted cliques [10] in a graph. A clique *C* is a subset of *V*, such that *G(C)* is complete sub-graph. The maximum weighted clique problem is finding for cliques with maximum weight, which are not proper subsets of other clique. Here the weighted Clique number is the total weight of the maximal clique:

$$\omega(G, W) = \max\{W(C)\colon C\ is\ clique\ of\ G\} \qquad (3)$$

The problem we want to solve is to *find the minimum number of maximal cliques to cover the entire graph* (such that $\bigcup_i C_i \triangleq G$). This is application of a known graph-based problem in literature, denoted as *Clique Covering Problem* [10]. In particular, a clique cover of G is denoted as $\theta(G)$, clique cover of K cardinality can be interpreted as partitioning the graph into K maximal cliques $C_{n_1}, C_{n_2}, \dots, C_{n_k}$, where $n_k, \forall k \in K$ is the size of the maximal clique C, such that $\sum_{k \in K} C_{n_k} = |V|$. This can be translated as finding the minimum K maximal weighted cliques or:

$$\theta^*(G) := arg_K \min\{\max_{\omega} \sum_{k=1}^{K} \omega(C_{n_k}, W)\} \quad (4)$$

B. Controller Selection Sub-Problem

The problem in this case is finding the node (or nodes) per clique, which has the maximum degree within each clique C. Note that, the slice requirements might also affect the decision on the controller. So, $Cs_{u,v}(s)$ is another factor which will need to be considered. The degree of a vertex $u \in C$ can be defined as $\deg(u) = \sum_{v \neq u} w_{u,v}, \forall u, v \in C$. The problem for each maximal clique $C_i \subseteq G$ can be formulated as:

$$\mathrm{u^*}_{\mathrm{C_i}}(\mathrm{s}) \coloneqq \operatorname*{argmax}_{\mathrm{u}} \sum_{v \neq u} w_{u,v} Cs_{u,v}(s),\ \forall u \in C_i,\ \forall s \in S \quad (5)$$

C. RRM Split Configuration Sub-Problem

Finally, as soon as the cluster and the controller is decided he optimal configuration to maximize performance, subject to constraints as imposed by RAN characteristics and slice requirements. Let $A = \{a_{\mathrm{u^*}_{\mathrm{C_i}}(\mathrm{s}),v,s} | a_{\mathrm{u^*}_{\mathrm{C_i}}(\mathrm{s}),v,s} \in \{0,1,2\}\}$ be the variable corresponding to the RRM Split decision (e.g. 0 for centralized, 1 for semi-distributed and 2 for distributed RRM). This sub-problem can be seen as an Integer Programming problem where for each controller-slave TRP link we need to select which is the optimal split.

$$max \sum_{s \in S} \sum_{C_i \in G} \sum_{v \in C_i} a_{\mathrm{u^*}_{\mathrm{C_i}}(\mathrm{s}),v,s}\, Cs_{\mathrm{u^*}_{\mathrm{C_i}}(\mathrm{s}),v}(s), \forall v \neq \mathrm{u^*}_{\mathrm{C_i}} \in \mathrm{C_i} \quad (6)$$

, subject to traffic demand, backhaul & access channel, maximum delay and slicing constraints.

IV. Solutions Framework

In this section, mechanisms are presented that decide on how to form the clusters, what is the role of the nodes in each cluster (e.g. RRM Controller or Slave TRP) and who is forming the clusters (e.g., NMS based on physical deployment, long term statistics of the load of eNBs and slice characteristics or KPIs). In central management entity (NMS), the formation of appropriate clusters is selected. Subsequently, the level of centralized RRM is decided for each access node, given the RAN limitations and the level of slice awareness. The level of centralization is translated as a flexible split of RRM functions, which can be slice-tailored and cell-specific. The heterogeneous split of RRM functions will provide new requirement for signalling between the TRPs. For example, as mentioned in state-of-the-art, by centralizing only slower RRM functions like Interference Management and Load Balancing, signalling should be exchanged for the resource restrictions and cell re-selections between the centralized and distributed nodes for the dynamic resource allocation. In case of having distributed allocation of IM and LB, e.g. due to slice requirements for fast IM, we need to exchange new messages regarding the dynamic resource restrictions in order to allow for centralized LB (taking into account and the other RRM Splits). In this section, we propose some low complexity heuristic approaches, with low signalling cost to solve these sub-problems.

A. Cluster & Controller Selection

As mentioned above, we try to find sets of feasible solutions (e.g. multiple maximal cliques) to form the clusters. From the set of maximal cliques, we aim to find nodes with the highest occurrences to become candidates for RRM Controllers. The proposed heuristic algorithm is presented below.

Step 1 Initially each TRP creates an adjacency matrix which encloses neighbour TRPs. This matrix is forwarded to NMS, where the graph is created and manipulated. The TRP-TRP entry is 0 where non-ideal BH exists or distance is higher than a threshold (or 1 otherwise).

Step 2 NMS then Creates a graph G (V,E) where V is the TRP set and E are the weighted edges as discussed above. The edges are defined from the adjacency matrices and take also into account the processing capabilities, load of TRPs and the per-slice weighting factor.

Step 3 For each slice, starting from a random node find complete sub-graphs with the maximum weight and remove them from the graph.

Step 4 Repeat **Step 3** for all TRPs

Step 5 Repeat **Step 2** for all network slices

Below, in Fig. 4, the flowchart is illustrated:

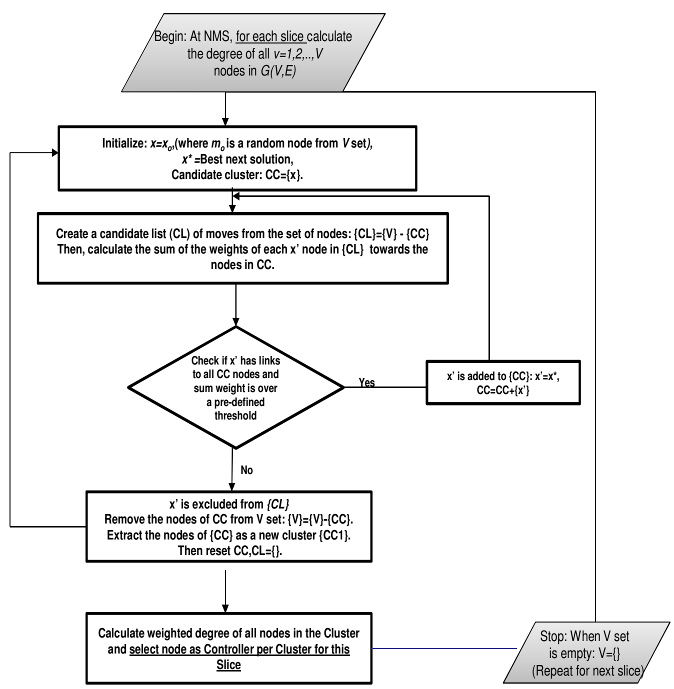

Figure 4 Flowchart for Controller and Cluster Selection

B. RRM Split Configuration

In this solution, we aim to find what the best split is for each controller-slave TRP pair based on the parameters as mentioned above. Here, by taking into account the BH constraint, the per slice preference on certain split and the load of the controller, we decide whether to use Split A (centralized), Split B (semi-centralized) or Split C (distributed). Below, in Fig. 5, we briefly show the flowchart for this selection.

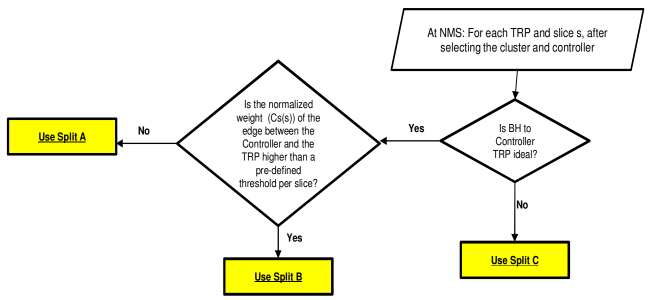

Figure 5 Flow-chart for RRM option configuration

An example system, where 2 clusters (complete sub-graphs) consisting of TRPs, which are selected as the cliques with the minimum cardinality, can be seen in Fig. 6. In this case, we select one controller per cluster and the RRM Split. In Cluster 1, we can observe that Slow RRM (e.g. Cell Selection) happens centrally, whereas the Dynamic RRM is performed in distributed way. In similar manner in Cluster 2, the Controller perform centralized RRM for some TRPs, whether one TRP can also perform dynamic RRM.

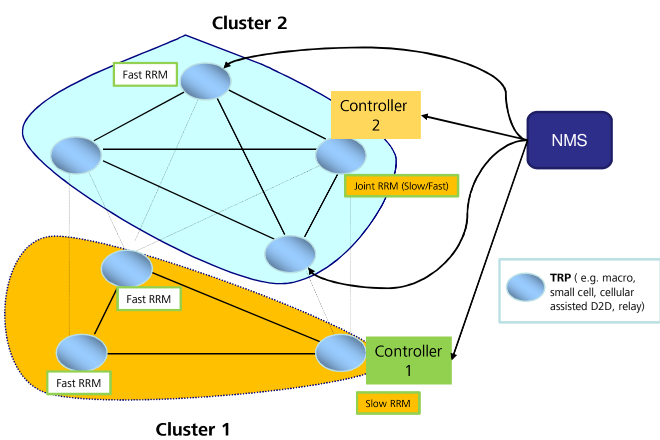

Figure 6 Exemplary scenario with 2 Clusters and 2 Controllers

V. Signaling Process and Evaluation

System level Simulations were performed to show the tradeoff between Centralized and Distributed Interference Management. The deployment is a Cluster of 9 TRPs using 3GPP LTE as baseline for our simulations (40 users uniformly distributed, 3GPP UMi channel [11], ideal BH). In case of Centralized RRM we perform Centralized CoMP (coherent JT), in case of Split B we perform only centralized eICIC and in case of Split C we perform single-cell scheduling in each TRP without interference management.

At first, NMS selects the RRM Controller and the Cluster Size the characteristics of the physical nodes (e.g., BS, (non) ideal BH link, available spectrum) in case of slice support the KPIs of every slice and also based on long term statistics for load, BH conditions per deployed slice. The NMS according to the previous data will decide on an Initial RRM Split. NMS configures each TRP as on its cluster membership, its operation mode as an RRM Controller (or simple member) and which initial RRM split will be used between the TRP and the RRM Controller.

Initially, an event (e.g. slice instantiation request) triggers the action from NMS to decide on how to form the clusters/who is forming the clusters (e.g., NMS based on physical deployment, long term statistics of the load of eNBs) and which RRM Split to be used. This can be configured by NMS. The selection of cluster, controller and RRM split for a certain slice, will divide the total set of access nodes into orthogonal TRP clusters. Thereafter, network will provide one or more access nodes as RRM Controller candidates based on the following parameters: TRP General Processing capabilities, Average load information, number of neighbouring TRPs with good / ideal BH, slice KPIs. Then, based on the slice requirements different RRM controller candidates can be mapped to different slices. In Fig. 7, we can also observe the message sequence chart for the process. In addition to the typical operation between the TRPs and NMS, which involves the feedback of measurement and long term statistics periodically, the Controller assignment message is forwarded from NMS to the Controller TRP and the Cluster member assignment is then forwarded to other TRPs. Here, two new messages can be defined for the controller assignment and cluster notification.

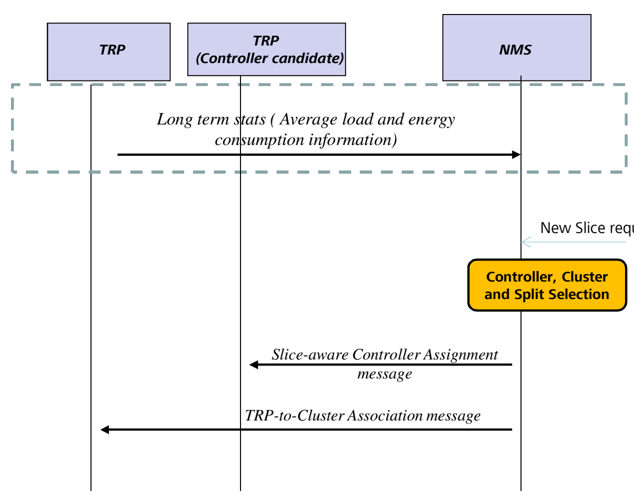

Figure 7 Message Sequence Chart for Controller and Cluster Selection

As mentioned above, for different slices we may have different requirements for spectral efficiency. For the example shown in a practical scenario, for URLLC more than 1b/s/Hz is acceptable level, while for eMBB more than 2.5b/s/Hz spectral efficiency is required. Thus, we select the level of centralization given these requirements and the interference levels (e..g for cell edge users we might need centralization to benefit from multi-connectivity at cell edges).

The per-TRP Spectral Efficiency for this particular simulation setup can be seen in Fig. 8. As we can observe from the CDF of spectral efficiency (SE), for MTC slice we do not need to centralize RRM, unless the users are near the cell-edge, since the SE KPI is fulfilled. On the other hand for eMBB the higher the centralization the higher gain we can achieve.

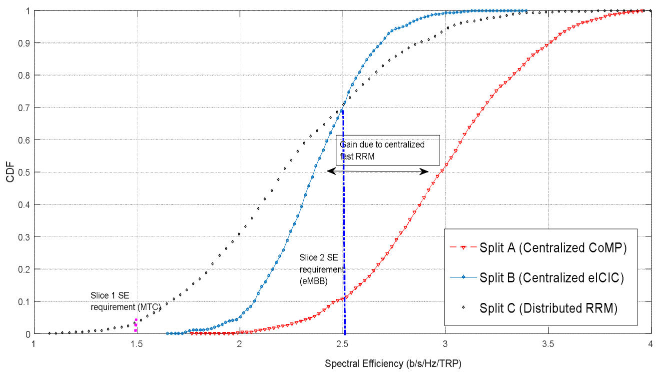

Figure 8 CDF of Spectral Efficiency – Comparison of different splits

One important aspect is the signalling overhead and the data feedback that needs to be forwarded per BH link (e.g. for JT CoMP in case of centralized fast RRM), which will also affect the selection of the controller and the split. The data exchange required as can be seen in Fig. 9 would be high for centralized solutions up to 3 TRPs in a cluster, but for large number of clusters centralized RRM will need less backhaul for the data exchange overhead. The computation of the BH required is discussed in [12]. For this calculation of BH rate requirement 600 Sub-carriers with 15KHz spacing were assumedwith 8-bit quantize-and-forward method. This shows that another factor that needs to be considered is whether we are able to offer certain centralization given the physical deployments and the density of access points. As can be seen in Fig. 9, in centralized CoMP (split A), the load increases linearly (2 x number of TRPs), which can be much lower for large clusters than distributed (split B). So, for very dense RAN, Split A might be more preferable than Split B in case that high SE (through JT CoMP) is required to achieve very high capacity gains (e.g. for eMBB); otherwise to achieve the same gains in distributed case we will need much higher BH rate requirements (for all point-to-point exchanges) as the cluster size increases.

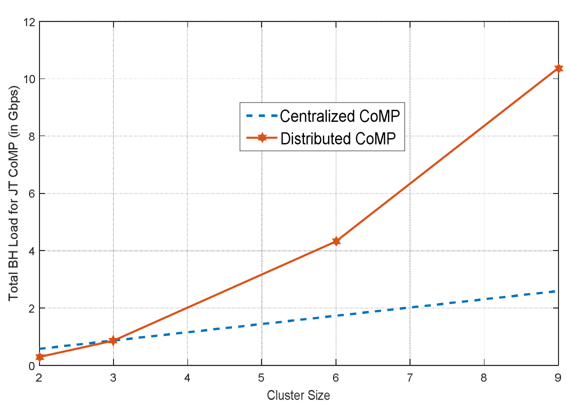

Figure 9 Backhaul Load requirement – Comparison for Centralized vs. Distributed CoMP

VI. CONCLUSION

In this paper, the problem of selecting a set of clusters controller from the RAN nodes in order to perform slice-aware RRM is discussed. In this context, we present the problem framework in a graph-based framework. A solution is provided for adaptively placing the RRM Controller in different TRPs taking into account the processing capabilities, the easy reach to the other TRPs, the load conditions and the slice requirements. This will provide 1) Cost Efficiency, since we do not need a dedicated entity to act as Controller per cluster but the controller might reside in different entities, 2) Adaptation to a dynamically changing RAN environment since the shape and density of the Clusters might change, 3) C-Plane delays will be expected to be lower since the distribution of RRM functionalities in multiple RRM Controllers will allow some control information to be locally exchanged, 4) Slice-awareness in RAN with the minimum impact on RAN design, since the only new information for the slices in RAN nodes will be the controller selection and the selected split.

ACKNOWLEDGMENT

Part of this work has been performed in the framework of the H2020 project METIS-II co-funded by the EU. Authors would like to acknowledge the contributions of their colleagues from METIS-II although the views expressed are those of the authors and do not necessarily represent the views of the METIS-II project.

REFERENCES

[1] Next Generation Mobile Networks (NGMN) Alliance, “5G White Paper”, Feb. 2015.

[2] 3GPP TR 38.913, “Study on Scenarios and Requirements for Next Generation Access Technologies”, March 2016.

[3] White Paper, *5G PPP Architecture Working Group - View on 5G Architecture,* July 2016

[4] Yousaf, Faqir Zarrar, and Tarik Taleb. "Fine-grained resource-aware virtual network function management for 5G carrier cloud." IEEE Network 30.2 (2016): 110-115.

[5] 3GPP TS 36.300: Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2, v. 13.3.0, April 2016.

[6] T. O. Olwal; K. Djouani; A. M. Kurien, "A Survey of Resource Management towards 5G Radio Access Networks," in *IEEE Communications Surveys & Tutorials* , vol.PP, no.99, pp.1-1

[7] Sabella *et al*., "RAN as a service: Challenges of designing a flexible RAN architecture in a cloud-based heterogeneous mobile network," *Future Network and Mobile Summit (FutureNetworkSummit), 2013*, Lisboa, 2013, pp. 1-8.

[8] D. H. Lee, K. W. Choi, W. S. Jeon and D. G. Jeong, "Two-Stage Semi-Distributed Resource Management for Device-to-Device Communication in Cellular Networks," in *IEEE Transactions on Wireless Communications*, vol. 13, no. 4, pp. 1908-1920, April 2014.

[9] D. W. K. Ng and R. Schober, "Resource Allocation and Scheduling in Multi-Cell OFDMA Systems with Decode-and-Forward Relaying," in *IEEE Transactions on Wireless Communications*, vol. 10, no. 7, pp. 2246-2258, July 2011.

[10] LM. Bomze, M. Budinich, P.M. Pardalos, and M. Pelillo, ”The Maximum Clique Problem,” Handbook of Combinatorial Optimization, Kluwer Academic, 1999.

[11] 3GPP TR 36.814, “Evolved Universal Terrestrial Radio Access (E-UTRA); Further advancements for E-UTRA physical layer aspects (Release9)”, V9.0.0, Mar. 2010.

[12] P. Marsch and G. Fettweis, Coordinated Multi-Point in Mobile Communications: From Theory to Practice. Cambridge, U.K.: Cambridge Univ. Press, 2011.