# Hierarchical Network Abstraction for HetNet Coordination

Sergio Lembo*, Junquan Deng*, Ragnar Freij-Hollanti*, Olav Tirkkonen* and Tao Chen†
*School of Electrical Engineering, Aalto University, Finland
†VTT Technical Research Centre, Finland

***Abstract*—We consider a user-centric network-level coordination architecture for 5G heterogeneous Radio Access Networks (RANs), based on RAN softwarization and a centralized coordination framework. We describe the RAN as a set of logical RAN entities, related to cells in a Heterogeneous Network (HetNet), under the control of a central coordination entity. This description allows the creation of Network Functions (NFs) with an abstracted view of the network. We describe a centralized coordination framework, and then develop a NF for Inter-Cell Interference Coordination (ICIC) in a 5G HetNet, optimizing the radio resource usage at network-level. We construct a Network Graph to abstract the problem of resource allocation and cell offloading, with the NF seeking for an optimal solution based on this abstraction. Simulations are performed in a HetNet scenario with a Tabu Search algorithm. Results show the feasibility of performing network-level coordination through a modular NF, with an abstracted view of the network.**

## I. Introduction

Fifth generation (5G) mobile networks will provide consistent user experience for a multitude of use cases supporting a mobile and connected society [1]. Apart from advances in new physical layer technologies, such as millimeter-wave and massive Multiple-Input-Multiple-Output (MIMO) technologies [2], disruptions in the logical structure of mobile networks are required. In [3], essential architectural requirements for 5G Radio Access Networks (RANs) were discussed. Among others, dual connectivity between LTE and 5G, multipoint transmission, control and user plane separation, as well as interfaces supporting inter-site scheduling coordination were identified as crucial 5G features.

In LTE-Advanced networks, Heterogeneous Networks (HetNets) with related Inter-Cell Interference Coordination (ICIC) [4], load balancing and mobility management problems, as well as development of MIMO technologies and concepts of Coordinated Multipoint (CoMP) transmissions, have increased the need for coordination of RAN nodes. Emerging 5G technologies such as dynamic Time-Division Duplexing (TDD) [5], Ultra-Dense Networks (UDN) and direct Device-to-Device communication (D2D) [6] further increase the need for tight coordination of 5G RAN nodes.

In the control and management planes, centralized architectures have been prevalent. In this context, different paradigms propose the use of a logically centralized coordination entity with plug-in Network Functions (NFs), or Applications, to control or manage the communication network. Software Defined Networks (SDN) proposes a logically centralized SDN controller in charge of the control plane, behaving as a network operating system. It is complemented by programmable SDN applications with an abstracted view of the network [7]. Network Function Virtualization (NFV) proposes a centralized Orchestrator [8]. Self Organizing Networks (SON), with centralized control [9], introduces the idea of a SON coordinator or controller, where SON functions are plugged-in. Here, the central controller collects reports about the Base Stations (BSs) and the users served, which are processed by a NF, linked through a suitable Application Programming Interface (API) [9].

Motivations for the design of a software-defined architecture are the possibility to attach modular NFs, minimize standardization time of new functions, simplify network management, and encourage innovation. Recently, the paradigm of SDN has been extended to mobile RANs [10][11]. However, while NFs for SDN are well defined, NFs for RAN still need to be developed. One of the challenges is the adoption of agnostic RAN interfaces, which need to be defined through proper network abstractions. In [10], the whole RAN is treated as a virtual base station, whereas in [11] hierarchies of SDN control with layers of abstraction were considered. In [6], options for integrating SDN with 5G networks with both Distributed-RAN and Cloud-RAN architectures, supporting UDN, massive-MIMO and other 5G technologies were considered. A hierarchy of partially distributed and partially centralized NFs was predicted for 5G, where sites of Macro Cells (MCs) would be locations for locally centralized RAN control functions, governing UDNs, D2D, and other local area subsystems.

Ultimate solutions for providing consistent user experience would be based on user-centric architectures, where network resources follow users. In the CoMP literature, dynamic cell clustering approaches have been widely discussed [12][13][14]. In [12][13], Coordination Areas (CA) of multiple cells were formed, where cells belong to multiple CAs, depending on the user population. In [14] virtual Baseband Units (BBUs) were discussed, serving users in intersections of cells. Resources in both cells are reserved to serve these users. Such solutions could be enabled by inter-site scheduling solutions called for in [3]. In [14], clustering of the users

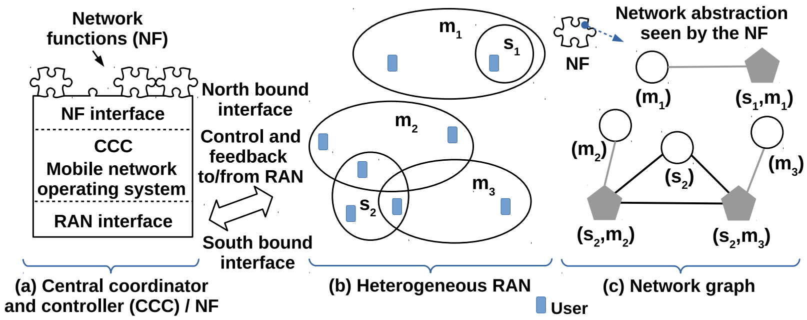



Fig. 1: (a) Central Coordinator and Controller, and Network Functions. (b) A heterogeneous network with three MCs, two SCs, and associated users. The SCs are used for offloading users of MCs. (c) A small network graph of logical entities derived from the network in Fig.(b). Circles: Single-cell entities. Pentagons: Multi-cell entities. Black lines: Conflict edges (LREs share at least one cell). Gray lines: weighted edges.

to virtual BBUs was seen to be the main inter-node problem.

In this paper, we consider a user-centric network-level coordination architecture for 5G heterogeneous RANs, based on RAN softwarization and a centralized coordination framework. First, we propose a logical description of the RAN by Logical RAN Entities (LREs), under the control of a central coordination entity, where each LRE is a set of cells in a HetNet. Such description allows higher flexibility in the coordination of networks that involve multiple tiers or multi-connectivity, and is suitable for creating NFs with an abstracted view of the network. Then, we describe a centralized coordination framework. Users are associated with the LREs. Based on user and RAN measurements, the LREs formulate NF-specific metrics. These metrics are communicated to a Central Coordinator and Controller entity (CCC), which determines the current state of the network based on the collected metrics, and communicates control information to the RAN nodes. The dynamically determined set of LREs, the user association principle and the metric calculation form a set of SDN-primitives. These enable user-centric centralized RAN coordination where the CCC coordinates the RAN nodes responsible for air-interface communication with the users. The state of the network related to the NF is abstracted and described by a Network Graph (NG), with nodes representing LREs. Finally, we develop a NF for ICIC, to demonstrate a concrete and working example of the proposed framework. The ICIC-NF optimizes radio resource usage at network-level. Describing the network in terms of multi-node LREs allows to solve disputes for resources among the cells. The network-level ICIC NF is realized based on a Tabu Search algorithm to find optimal muting ratios. The NF is plugged in into a system level simulator. Simulations are performed in a HetNet scenario to observe the feasibility of adopting network-level coordination through a centralized framework with a modular NF.

## II. Network Abstraction with LREs

A CCC with plug-in NFs coordinates a HetNet RAN, Fig. 1(a). The southbound interface of the CCC communicates with the RAN, and the northbound interface with various NF applications, c.f. [11]. The CCC aggregates abstracted information gathered from the RAN, and with the NF determines the control signaling to the RAN.

The HetNet consists of a set of cells, with a Small Cell (SC) tier deployed under the umbrella of a macro network, and a set of users. Depending on the NF, the users may be instantaneously active users, idle mode camping users, or entire historical user populations. We propose a NF-specific user association function, which associates a user with a LRE, consisting of a cell or a set of multiple cells. An example of such logical entities and of associating users to them is depicted in Fig. 1(b)(c). The set of logical entities is thus a subset of the power set of the cells in the RAN.

The RAN nodes create software agents to govern the service of the users associated with a LRE. These agents are created in real-time controllers within the RAN nodes, which are responsible for user and control plane communication with the users over the air interface. They may be located in, e.g. Cloud-RANs, or realized in a distributed manner. LREs communicate with the CCC via the southbound interface NF-specific metric calculations, Fig. 1(a). These convey an abstraction of the state of the RAN protocols. Depending on the NF, various properties of the physical, and medium access layers are aggregated to metrics indicating either the state of the LRE, or the interaction between LREs. The CCC informs to the LREs its decisions about the NF, as well as specifics of metric calculations, user association principles, etc.

Based on the collected information, the CCC can construct a NG describing the state of the network related to the NF. The NG may be weighted both at the vertices and at the edges, with vector valued weights. The vertex weights represent vector states, and the edge weights interactions between entities. An example is depicted in Figure 1(c). There are single-cell entities and multi-cell entities. Certain operations are only possible between entities that do not share a cell. Accordingly, there may be conflict edges between LREs which share at least one cell. Between LREs that do not share cells, there may be weighted edges, indicating, e.g., interference couplings.

A logical description of the network in terms of LREs is beneficial for all NFs that involve pairs or larger sets of cells. Thus for dual connectivity, control/user plane separation, load balancing and mobility management, LREs responsible for pairs of cells can be used. For dynamic TDD, ICIC, network-controlled D2D, multi-point transmission and CoMP, some users may benefit from entities consisting of more than two cells. Mobility management is a representative of a NF for which information gathered from historical user populations may be of great value.

## III. Network-level ICIC: System Model

We consider a HetNet, Fig. 2(a), with users $u$ in the sets $\mathcal{U}_c$, with cardinality $U_c = |\mathcal{U}_c|$, where $c$ denotes the cell to which $u$ is associated. Here $c$ belongs to a set of MCs $\mathcal{C}_m$, or SCs $\mathcal{C}_s$, with cells $m \in \mathcal{C}_m$, and $s \in \mathcal{C}_s$. The set of all the cells is $\mathcal{C} = \{\mathcal{C}_m \cup \mathcal{C}_s\}$. The total number of users is $U = |\bigcup_{c\in\mathcal{C}} \mathcal{U}_c|$. Each cell has a BS transmitting in downlink. Let $\mathcal{K}$ be the set of muting combinations considered for MCs, with $K = |\mathcal{K}|$. In the general case there are $K = 2^M$ muting combinations, where $M = |\mathcal{C}_m|$. The muting configurations are indicated by vectors $\mathbf{k} \in \{0,1\}^M$, which hold the *on/off* status of each $m$, meaning that when $m$ is *off*, it mutes transmission, and $k_m = 0$. Muting $m$ increases the Signal-to-Interference-plus-Noise Ratio (SINR) of users offloaded to the Cell Range Extension (CRE)-region of SCs. CRE is proposed by 3GPP to ensure that a sufficient amount of user traffic is offloaded to SCs by using a cell association offset to the received power in the SC [4]. There is multiplexing gain in muting, according to the number of SCs that benefit from the muting.

The general problem is to find the scheduling weights of the resources allocated to the users, such that a network-level utility is maximized. Resources can be in time or frequency, are orthogonal among users in the same cell, and are assumed non discrete. The optimization objective is,

$$\mathcal{W}^* = \arg\max_{\mathcal{W}} \sum_{c\in\mathcal{C}} \sum_{u\in\mathcal{U}_c} g_u, \tag{1}$$

where $\mathcal{W}$ is the set of scheduling weights, and the term to maximize is the sum of the utilities $g_u$ of all the users in the system. Here we consider a proportionally fair utility function for the allocation of resources among the users [15], so that

$$g_u = \log\left(\sum_{\mathbf{k}\in\mathcal{K}} w_{u,\mathbf{k}} x_{u,\mathbf{k}},\right) \tag{2}$$

where $w_{u,\mathbf{k}}$ and $x_{u,\mathbf{k}}$ are the scheduling weight and Spectral Efficiency (SE) of a user $u$ in the combination $\mathbf{k}$, respectively. The sum of resources of all the users in a cell, for each $\mathbf{k}$ is

$$v_{\mathbf{k},c} = \sum_{u\in\mathcal{U}_c} w_{u,\mathbf{k}} \quad \text{for } c \in \mathcal{C}, \text{ with } w_{u,\mathbf{k}} \geq 0, \tag{3}$$

and the sum of these over all the muting combinations is constrained to the total amount of resources normalized to one,

$$\sum_{\mathbf{k}\in\mathcal{K}} v_{\mathbf{k},c} \leq 1 \quad \text{for } c \in \mathcal{C}. \tag{4}$$

In addition, the slices of resources for each $\mathbf{k}$ are the same for all the cells,

$$v_{\mathbf{k},c_i} = v_{\mathbf{k},c_j} \quad \text{for } c_i,\ c_j \in \mathcal{C}, \text{ and } \mathbf{k} \in \mathcal{K}. \tag{5}$$

When there is a transmission from cell $c$, it is received at user $u$ with the power $R_u^{(c)}$. The SE of a user in a combination $\mathbf{k}$ is obtained from the SINR. For analytical purposes we consider,

$$x_{u,\mathbf{k}} = \log_2\left(1 + SINR_{u,\mathbf{k}}\right) \tag{6}$$

where the SINR depends of the cell association of $u$,

$$SINR_{u,\mathbf{k}} = f\left(R_{u,\mathbf{k}}, I_{u,\mathbf{k}}\right) \tag{7}$$

Here, $f(\cdot)$ is the SINR for the antenna arrangement and diversity in use, as function of the received signal at $u$ from the serving BS, and the interference powers. For a user served by MC $m$, the serving BS power is $R_{u,\mathbf{k}} = k_m R_u^{(m)}$, whereas for a user served by SC $c$, we have $R_{u,\mathbf{k}} = R_u^{(s)}$. The interference powers are

$$I_{u,\mathbf{k}} = \sum_{m\in\mathcal{C}_m} k_m R_u^{(m)} + \sum_{s\in\mathcal{C}_s} R_u^{(s)} + N_0 - R_{u,\mathbf{k}} \tag{8}$$

Where, $N_0$ is the noise power. When $m$ is muting, the received power from $m$ is switched off according to $k_m$.

In the centralized coordination framework discussed here, we could consider gathering the SEs of all the users in the CCC and solve the optimization problem stated by (1), (3), (4) and (5). The optimization formulation (1) is convex. However, the cardinality of the set of muting combinations is exponential in the number of cells. Accordingly, the problem can be proven to be NP-hard, following the approach of [16]. Even for a moderate network, the computing effort and signaling overhead is big. Therefore, we need to resort to some abstraction to simplify the problem and produce a cost efficient solution. This is addressed in the next section, using the framework of section II.

## IV. Network-level ICIC: Network Function

In this section we develop a specific NF for network-level ICIC in a HetNet based on a logical description of the network in terms of LREs. ICIC is an important NF for co-channel 5G HetNets since it can significantly improve the SINR performance of the users located at the intersection of two or more cells. In a user-centric deployment, a HetNet, with multiple BBUs and a cluster of Remote Radio Heads (RRHs) attached to each BBU, would be deployed according to the spatial/temporal distributions of the users. The dense and irregular deployment of RRHs forming SCs, together with co-channel MCs, would introduce complicated interference interactions between these RAN elements. Load balancing in such a HetNet becomes difficult when the network dimension increases. Serving users in the CRE areas requires RAN coordination at network-level when the SC receives strong interference from more than one MC. For a given CRE offset, users fall into one of

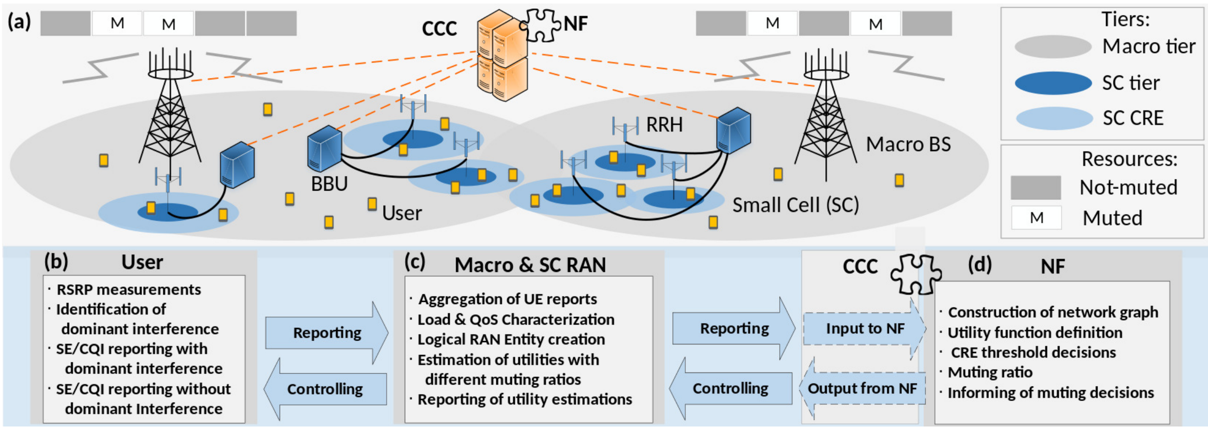



Fig. 2: Hierarchical Control Framework for Network-Controlled load balancing and ICIC.

three sets: SC users, MC users and CRE users. In the centralized coordination framework discussed here for network-level ICIC, it is natural to use LREs to govern these user sets.

### A. Hierarchical Control Framework for ICIC

Fig. 2(a) shows a hierarchical control framework of network-level ICIC. The physical RAN entities include BBUs which are responsible for multiple SCs $s$, and MCs $m$. The users, Fig. 2(b), are responsible for measuring Reference Signal Received Power (RSRP), identifying dominant interferences, and reporting a metric based on average SEs, or Channel Quality Indicators (CQIs) with suitable granularity, with and without the dominant interferences. The RAN nodes, Fig. 2(c), are responsible for creating LREs, aggregating user estimates, and reporting metrics to a CCC, as shown in the figure. For example, an utility metric of the users in a LRE, based on SE or CQI, can be computed for a given amount of resources for posterior reporting to the CCC. The LREs are created inside the RAN node, under control of the CCC, and form the nodes of a NG for centralized RAN coordination.

The CCC with the ICIC NF, Fig. 2(d), is in charge of constructing the NG and implementing a suitable logic for the network coordination. For a given NG, we say that a resource allocation pattern is *conflict-free* if two LREs that are connected by an edge are never assigned the same resource. Then, two different kinds of interactions between LREs generate edges in the NG:

- Conflict edge: for set of LREs that share the same physical RAN node (SC), the resource usage pattern is orthogonal between the different LREs.
- Weighted edge: if users of one LRE receive dominant interferences from another LRE, these LREs should avoid using the same radio resources. Here, MCs $m$ that strongly interfere some users in $s$, are muted on the resources on which these users are served.

The construction of the NG allows us to consider these two interactions simultaneously. With this definition, a conflict-free resource allocation, where each LRE is served by one network resource, is known as a *graph coloring* in graph theory. Here, we want to assign some resources to each LRE, in a conflict-free way. Let $\mathcal{L}$ be the set of LREs in the network and let $\mathcal{R}$ be the set of resources, with power set $P(\mathcal{R})$. The natural mathematical abstraction of the problem is thus to assign a set-valued function $\Phi : \mathcal{L} \to P(\mathcal{R})$, assigning a set of resources to each LRE, such that $\Phi(a) \cap \Phi(b) = \emptyset$ if $ab$ is an edge in the NG. This is known as a *multicoloring* problem.

### B. NG and creation of LREs

A NG is constructed to manage the interference situation produced by the strong interference that macro BSs exert into SC users in the CRE region. Then, LREs can be created and removed dynamically according to need, for example after cell-user association. Each user $u$ associated with a MC $m$ is further associated with a LRE $(m)$. Each user $u$ associated with a SC $s$, and with the presence of a set $\mathcal{M}$ of MC strong interferers, is associated with the LRE $(s, \mathcal{M})$. If $\mathcal{M}$ is the empty set, then we denote $(s, \mathcal{M})$ by $(s)$, and if $\mathcal{M} = \{m\}$ is a singleton, then we denote $(s, \mathcal{M})$ by $(s, m)$.

Resources are interchangeable among LREs respecting conflicts. In other words, the utility function is invariant under permutations of $\mathcal{R}$. In our abstraction, we will actually use the stronger assumption that the utility function only depends on the *amount of resources* assigned to each LRE, as long as the resource allocation is conflict-free in the sense of the previous section. In each MC, resources can be muted with an independent muting ratio. Resources for local scheduling in each LRE are assumed infinitely sub-divisible among the users in the LRE, which is an approximation of long time duration or wide band.

Hereafter, we assume that reports of SEs are available at the RAN node, or that a suitable approximation can be obtained from CQIs. Moreover, it is assumed that the SEs come from an aggregation of the SEs reported by the users, for example the average SE, which helps to abstract the fine details of the state of the channel. Aligned with [11], the proposal is to leave real time decisions of fine granularity to the local scheduler in the RAN node, and let the CCC specify the share of resources assigned to the LRE, on a time scale of seconds.

RAN network entities can compute a utility for users, and a sum utility for LREs, from the SEs reported by the users. Here, for concreteness, the sum utility proposed for a LRE is computed assuming a local proportional fair distribution of resources [15]. In average, each user receives $1/p$ resources, where $p$ is the number of users in the LRE. Then, the sum-utility in the LRE for a share of $w$ resources assigned to the LRE is

$$U_{\mathrm{LRE}} = p(\log(w) - \log(p)) + \sum_{i=1}^{p} \log(x_i), \qquad (9)$$

where $x_i$ is the SE of user $i$, calculated from the SINR. For LREs $(m)$ and $(s)$ the SE is calculated with (6), assuming a muting combination $\mathbf{k}$ where all the BS are transmitting. For LREs $(s, \mathcal{M})$ the SE is calculated with (6), assuming a muting combination $\mathbf{k}$ where only the strong interferers, determined with a threshold, are muted. For reallocating resources, the CCC only needs to know the number of users in the LRE and, from (9), the aggregation of SEs $\sum \log(x_i)$, thus reducing the overhead in signaling, in contrast to the required to solve problem (1).

An adjacency matrix is created which describes conflicts between LREs. Edges are established between LREs that have at least one common member. For example, an edge between $(s)$ and $(s, m)$, indicates that cell $s$ serves users in $(s)$ and in $(s, m)$ with different resources. Also, an edge between $(s, m)$ and $(m)$, indicates that $m$ mutes in the resources given to $(s, m)$.

### C. Optimization Algorithm

A centralized multicoloring is performed. As the information is centralized, a global optimization problem can be constructed for resource allocation. In Section III we mentioned the unfeasibility of solving (1). Here we combine the NG with a Tabu Search (TS) technique [17], in which the algorithm progresses toward a maximum system utility by evaluating moves in the NG. TS has shown positive performance for this kind of problems [18]. The implementation is summarized in Algorithm 1. Function $f_{\text{AdjacencyMatrix}}(.)$ returns the LREs that have conflicts with the selected LRE from the Adjacency Matrix. Function $f_{\text{Moves}}(.,.)$ returns the candidate moves for a given LRE, these are the moves to explore neighboring solutions around the LRE. It is noted that the move is initiated from one LRE, but due to the moves and constraints there are many LREs participating in the move. Then, $f_{\text{Constraints}}(.,.)$ evaluates constrains as described below. The objective function to be maximized is the network utility $f_{\text{Ut}}(.,.)$. The move is accepted if it maximizes the best utility known at that moment, then the decision is communicated to the participating LREs to update the allocation of resources. The algorithm can iterate until a completion criterion is met, however, as the network is not static, a dynamic algorithm will iterate indefinitely, aiming to have a long term convergence.

**Algorithm 1** Tabu Search in Network Graph

```
TabuList ← [], BestUtility ← −∞
while completion criterion not met do
    SelectedLRE ← Select a LRE at random
    Conflicts ← f_AdjacencyMatrix(SelectedLRE)
    MovesList ← f_Moves(SelectedLRE)
    for CandMov in MovesList do
        BestMov ← null
        if f_Ut(CandMov,Conflicts) > f_Ut(BestMov,Conflicts) and
        CandMov ∉ TabuList and
        f_Constraints(SelectedLRE,Conflicts) respected then
            BestMov ← CandMov
        end if
    end for
    SelectedCandidateMove ← BestMov
    if f_Ut(SelectedCandidateMove) > BestUtility then
        Accept SelectedCandidateMove
        BestUtility ← f_Ut(SelectedCandidateMove)
    end if
    update TabuList, remove older entry
end while
```

### D. Moves and Constraints

The resource allocation problem is a multicoloring problem of the graph, in the sense that each node should be assigned a set of resources, such that the sets assigned to nodes that share an edge do not overlap. However, in this paper, we solve a relaxation of this problem, where instead of assigning sets of resources to the LREs, we assign a number of resources. The constraint that neighboring nodes should receive non-overlapping sets, is then replaced by a condition for each clique in the graph. Recall that a clique in a graph is a collection of nodes such that there is an edge between every pair. We will first present the constraints for our optimization problem, and then introduce the allowed moves in the algorithm.

*1) Constraints:* The constraints are given by different types of cliques in the graph, as follows.

A) Constraint coming from weighted edges: For a given SC $s$ and a given MC $m$, the LREs $(s, \mathcal{M})$ with $m \in \mathcal{M}$ can only be served by resources on which $m$ is muting. Thus, we have, $\sum_{\mathcal{M} \ni m} \phi(s, \mathcal{M}) + \phi(m) \leq 1$.

B) Constraint coming from conflict edges: For any SC, we have, $\sum_{\mathcal{M}} \phi(s, \mathcal{M}) + \phi(s) \leq 1$ (maximal number of resources used in a SC).

TABLE I: Simulation Parameters

| Description | Setting |
|---|---|
| Scenario 1<br>Scenario 2<br>Scenario 3 | $M$= 6 MCs, 4 SC per MC, (total: 30 cells)<br>$M$= 6 MCs, 6 SC per MC, (total: 42 cells)<br>$M$=12 MCs, 4 SC per MC, (total: 60 cells) |
| CRE | 20 dB |
| User distribution | $U_c$ = 60 users per MC.<br>$pU_c$ uniformly in MC<br>$(1-p)U_c$ uniformly in SCs of 50m<br>$p$ uniform in $[0, 1]$. |
| Macro BS inter-site dist. | 1732 m |
| MC / SC Tx Power | 43 / 30 dBm |
| User noise figure | 9 dB |
| Thermal noise level | -174 dBm/Hz |
| Path loss model MC/SC | UMa/UMi [20, Table B.1.2.1-1] |
| Shadow fading | Std. Dev. 8 dB |
| SINR to rate mapping | Shannon rate formula |

C) A LRE with no resources cannot decrement resources, and a LRE with the maximum available resources cannot increment resources.

*2) Resource-moves:* These aim to optimize the muting ratios for a given cell association, resulting in higher data rates. Moves started at LREs $(m)$ and $(s, \mathcal{M})$:

I LRE $(m)$ mutes: At most one $(s, \mathcal{M})$ per SC $s$ increments resources if the other $(m)$s with edges to $(s, \mathcal{M})$ are muted enough, consistent with the criterion for a weighted edge, respective $(s)$s decrement resources.

II LRE $(m)$ un-mutes: One $(s, \mathcal{M})$ per SC decrements resources, respective $(s)$s increment resources.

III LRE $(s, \mathcal{M})$ increments resources: $(m)$s with edges to $(s, \mathcal{M})$ are muted, $(s)$ decrement resources.

IV LRE $(s, \mathcal{M})$ decrements resources: $(m)$s with edges to $(s, \mathcal{M})$ are un-muted, if other $(s, \mathcal{M})$s with edges to $(m)$ are using an amount of muted resources as the total to remain muted, $(s)$ increments resources.

Function $f_{\text{Moves}}(.,.)$ returns moves I and II for a LRE $(m)$, and III and IV for a LRE $(s, \mathcal{M})$.

## V. Simulation and Results

To gather experience in the considered methodology we develop the ICIC NF described in section IV, and plug it into a custom made system level simulator through a CCC, which centralizes the information of the network. We simulate a HetNet consisting of MC and SC tiers. We assume a saturated system transmitting in downlink with infinite buffer model, no mobility, and non-quantized SINR information.

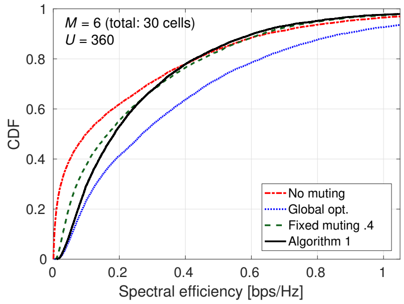
The layout consists of a grid of BSs with three MC sectors per BS site. Each MC operates in a SISO mode with sectorially divided antennas. Each cell has a dedicated scheduler. SCs are uniformly distributed within the MC coverage area, adopting the model of [19], assuming one SC per cluster, and one BBU per SC.

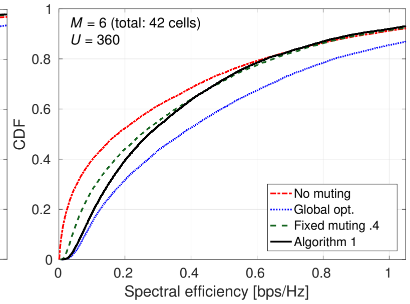
Simulations are executed in a number of network instances. In each simulation instance, BSs and users are distributed according to Table I. The random positioning of SCs and users is based on a Monte Carlo simulation, contemplating the evaluation of the proposed approach for different network settings. Users are associated to the cell which has the higher received power considering a CRE bias applied to the SCs. SINR for each user is calculated for the case that all the MC-BS transmit, and for the case that the strong interfering MC-BS are muted, depending the LRE-association. For LRE-association, MC users are assigned to LREs $(m)$. Then, SC users considered to be victim of strong interference are determined by comparing their SINR to a threshold set to 0 dB, and then assigned to $(s, \mathcal{M})$, and the remaining users to $(s)$. Strong interferers in $\mathcal{M}$ are determined by calculating the Signal-to-Interference Ratio, for the power $S$ of the signal of interest and the interference $I$ of an individual interferer, and comparing it to a second threshold of 40 dB.

Simulations are executed, with each RAN node, here BSs, reporting to the CCC aggregated SEs, and the number of users per LRE, as described for (9). The CCC operates with Alg. 1 in the NG, and reports to the MC-BSs the corresponding muting ratios, and to the SCs the $M$ muting ratios obtained. Then each cell determines locally the scheduling weights for the users[1]. The more information is extracted from the solution of Alg. 1, the better the coordination in the allocation of resources, at the cost of more signaling and complexity.

Scenarios 1, 2, and 3 are considered, with 30, 42 and 60 cells respectively (Table I). Figs. 3 to 5 show the CDF of average user rate, simulated in 30 network instances. Performance is evaluated in terms of SE, i.e. user rate per Hz of system bandwidth. We compare the performance of Alg. 1 in the ICIC-NF against the cases that there is no ICIC muting, fixed muting in all the network, and the optimal solution of (1). We use a solver to solve (1) in scenarios 1 and 2, with $10^5$ function evaluations [21]. We cannot solve (1) for scenario 3, due that the matrix to hold variables and constraints is of size $U_c M 2^M \times 2 U_c M 2^M$(126TB) [21]. Fixed muting with muting ratios .4 and .5 is considered as reference, the ratios were calculated counting victim users according to [22]. In the figures 3 to 5 we observe gain in the low percentiles and a trade-off of loss in high percentiles against no-ICIC and fixed muting. Table II shows that there is almost no gain in the average user rate, but from 49% to 77% gain per user in the 5th. percentile, compared to fixed muting. Solution of (1) is better, at the cost that we cannot solve the problem from scenario 3. A small greediness is observed for Alg. 1 against (1), however $\mathcal{K}$ has $K = 2^M$ versus $K = M + 1$ muting patterns in (1) and Alg. 1, respectively. In Fig. 5 we observe that Alg. 1 presents the same trend as in Fig. 3,4, suggesting the same performance regardless the size of the network. We used $2 \cdot 10^4$ iterations in Alg. 1, and

[1]The SCs are capable to construct $K = M + 1$ slices, equivalent to $v_{\mathbf{k},c}$ in (3). In each SC, $N_v$ victim users in $(s, \mathcal{M})$s, and $N_n$ non-victim users in $(s)$ are calculated. The scheduler allocates in each slice $w_{u,\mathbf{k}} = v_{\mathbf{k},c}/N_v$ resources to each victim user, until filling $N_v/(N_v + N_n)$ of the total resources, filling the $\mathbf{k}$ of all-muted MCs first. The remaining resources, $N_n/(N_v + N_n)$, are allocated in $w_{u,\mathbf{k}} = v_{\mathbf{k},c}/N_n$ fractions to the non-victim users, considering the appropriate filling of the $v_{\mathbf{k},c}$ in the boundary $N_v/(N_v + N_n)$.



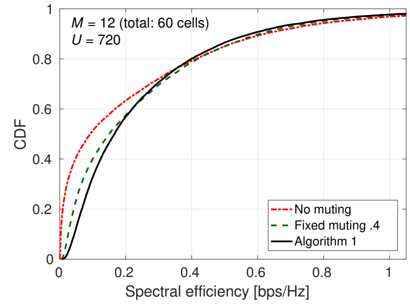




Fig. 3: Scen. 1. $M = 6$ MCs, 30 cells. Fig. 4: Scen. 2. $M = 6$ MCs, 42 cells. Fig. 5: Scen. 3. $M = 12$ MCs, 60 cells.

TABLE II: Gains of ICIC-NF with Alg. 1

| | | Scen. 1 | Scen. 2 | Scen. 3 |
|---|---|---|---|---|
| Mean | -no ICIC | 13% | 13% | 11% |
| rate gain ratio | -fixed muting 0.4 | 2% | 3% | 1% |
| of ICIC-NF over | -fixed muting 0.5 | 0% | 0 % | 0% |
| 5th. percentile | -no ICIC | 1879% | 1222% | 1787% |
| rate gain ratio | -fixed muting 0.4 | 77% | 74% | 77% |
| of ICIC-NF over | -fixed muting 0.5 | 50% | 49% | 50% |

tabu-list of size $M$, which gives a feasible computing time to update the network regularly, and with reduced signaling.

## VI. Conclusion

We have considered a user-centric, and centralized, network-level coordination framework for 5G RANs, based on RAN softwarization. We proposed a logical description of the RAN by logical RAN entities, under the control of a Central Coordination and Control entity in a Heterogeneous Network. RAN nodes construct Network Function specific Logical RAN Entities in a user-centric fashion. Information about the service that the LREs can provide to the users for different states of the NF, is communicated to the CCC, which decides on the network level state based on information aggregated from the whole network. The state of the network related to the NF was abstracted and described by a NG.

A NF for ICIC was considered to demonstrate a concrete and working example of the proposed framework, and its potential gains. The ICIC-NF optimizes muting ratios at network-level, it was implemented with a Tabu Search algorithm, seeking for an optimal solution through moves in the graph. Simulations performed in a HetNet scenario show that the abstracted view of the network in the proposed framework is suitable to improve the performance of the network. Alg. 1 can work with a dynamically changing network, and its performance is reflected by evaluating different network instances, with changing number of LREs, and different scenarios.


## Acknowledgment

This work has been carried out in the framework of the H2020 project ICT-671639 COHERENT, which is funded by the European Union.


## References


[1] R. El Hattachi and J. Erfanian (ed.), "5G white paper," NGMN Alliance, Tech. Rep., Feb. 2015.
[2] F. Boccardi, R. W. Heath Jr., A. Lozano, T. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Comm. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.
[3] 3GPP, "Study on scenarios and requirements for next generation access technologies," Tech. Rep. TR 38.913 v0.3.0, 2016.
[4] 3GPP, "Evolved universal terrestrial radio access (E-UTRA) and evolved universal terrestrial radio access network (E-UTRAN); overall description (R. 13)," Tech. Rep. TS 36.300; v13.3.0, 2016.
[5] P. Mogensen and K. Pajukoski & al., "5G small cell optimized radio design," in *Proc. IEEE GLOBECOM Wkshp,*, Dec. 2013, pp. 111–116.
[6] H. Droste & al., "The METIS 5G architecture," in *Proc. IEEE VTC* Spring, May 2015, pp. 1–5.
[7] ONF, "SDN architecture overview V.1.1," *ONF TR-504*, 2014.
[8] ETSI, "Network functions virtualisation (NFV); architectural framework," *ETSI TR, GS NFV 002*, Dec. 2014.
[9] M. Y. Arslan & al., "Software-defined networking in cellular radio access networks: potential and challenges," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 150–156, Jan. 2015.
[10] A. Gudipati & al., "SoftRAN: Software defined radio access network," in *In Proc. ACM SIGCOMM Wkshp. Hot Topics in Softw. Def. Netw.*, Aug. 2013, pp. 25–30.
[11] T. Chen, H. Zhang, X. Chen, and O. Tirkkonen, "Softmobile: Control evolution for future heterogeneous mobile networks," *IEEE Wireless Commun.*, vol. 21, no. 6, pp. 70–78, Dec. 2014.
[12] W. Mennerich and W. Zirwas, "User centric coordinated multi point transmission," in *Proc. IEEE VTC* Fall, Sep. 2010.
[13] S. Qin, C. Yu, O. Tirkkonen, and M. Alava, "Distributed coordination area formation in coordinated multi-point transmission," in *Proc. IEEE GLOBECOM*, Dec. 2011, pp. 1–5.
[14] F. Boccardi, O. Aydin, U. Doetsch, T. Fahldieck, and H.-P. Mayer, "User-centric architectures: Enabling CoMP via hardware virtualization," in *Proc. IEEE PIMRC*, Sep. 2012, pp. 191–196.
[15] J. Mo and J. Walrand, "Fair end-to-end window-based congestion control," *IEEE/ACM Trans. Netw.*, vol. 8, no. 5, pp. 556–567, Oct. 2000.
[16] L. Lei & al., "Optimal cell clustering and activation for energy saving in load-coupled wireless networks," *IEEE Trans. on Wireless Comm.*, vol. 14, no. 11, pp. 6150–6163, Nov 2015.
[17] F. Glover & al., "A user's guide to tabu search," *Annals of Operations Research*, vol. 41, no. 1, pp. 3–28, Mar. 1993.
[18] A. Hertz and D. Werra, "Using tabu search techniques for graph coloring," *Computing*, vol. 39, no. 4, pp. 345–351, 1987.
[19] 3GPP, "Small cell enhancements for E-UTRA and E-UTRAN - physical layer aspects (Release 12)," Tech. Rep. TS 36.872 v12.1.0, 2013.
[20] 3GPP, "Further advancements for E-UTRA physical layer aspects (Release 9)," Tech. Rep. TS 36.814; v9.0.0, 2010.
[21] Mathworks, "Global optimization toolbox: User's guide."
[22] S. Lembo, P. Lunden, O. Tirkkonen, and K. Valkealahti, "Optimal muting ratio for enhanced inter-cell interference coordination (eICIC) in HetNets," in *Proc. IEEE Int. Conf. Comm.*, Jun. 2013.