Universidad Carlos III de Madrid

Master on Telematics Engineering

Academic Course 2015-2016

Master Thesis

"Optimization of Resource Allocation Networks shared by multiple operators"

Author: Pablo Caballero Garcés

Director: Albert Banchs Roca

Leganés, 24th of January 2015

Keywords: **IMT-A, 3GPP LTE, Multi-tenancy, RAN Sharing, Fairness, Multi-operator.**

Summary: The traditional model of single ownership of the mobile network infrastructure is being challenged by the forecasted mobile data tsunami. In this context, the sharing of network infrastructure among operators has emerged as a way to ensure operators' future cost competitiveness. While the elementary concepts related to passive sharing are already exploited, active sharing is raising in importance. The work presented in this paper presents a solution for active RAN sharing in new generation cellular networks that takes into account inter and intra operator fairness and has desirable properties from base station and user association perspectives. An offline centralized approximation algorithm that provides theoretical performance guarantees is proposed. Making use of the available real-time neighbour information we designed an approximated distributed algorithm with an $\sum_{i=1}^{N} w_i \log(e)$ additive ratio. Finally, we confirm the theoretical results by a set of performance evaluations in a LTE-Advanced system-level simulator.

Resource Allocation for elastic traffic in mobile networks shared by multiple operators

Pablo Caballero Garcés*†
*IMDEA Networks Institute, Madrid, Spain †University Carlos III of Madrid, Spain

***Abstract*—The traditional model of single ownership of the mobile network infrastructure is being challenged by the forecasted mobile data tsunami. In this context, the sharing of network infrastructure among operators has emerged as a way to ensure operators' future cost competitiveness. While the elementary concepts related to passive sharing are already exploited, active sharing is raising in importance. The work presented in this paper presents a solution for active RAN sharing in new generation cellular networks that takes into account inter and intra operator fairness and has desirable properties from base station and user association perspectives. An offline centralized approximation algorithm that provides theoretical performance guarantees is proposed. Making use of the available real-time neighbour information we designed an approximated distributed algorithm with an $\sum_{i=1}^{N} w_i \log(e)$ additive ratio. Finally, we confirm the theoretical results by a set of performance evaluations in a LTE-Advanced system-level simulator.**

***Index Terms*—IMT-A, 3GPP LTE, Multi-tenancy, RAN Sharing, Fairness, Multi-operator.**

I. Introduction

Mobile networks are a key element of today's society, enabling communication, access and information sharing. All forecasts agree in predicting that the demand for capacity will grow exponentially over the next years, mainly due to video services. However, as cellular networks move from being voice-centric to data-centric, operators' revenues are not expected to be able to keep pace with the predicted increase in traffic volume. Such pressure on operators' return on investment has pushed research efforts towards achieving more cost-efficient mobile network solutions. In this context, network sharing has emerged as a key business model for reducing deployment and operational costs (CAPEX and OPEX).

Network sharing solutions are already available, standardized [2] and partially used in some mobile carrier networks. These solutions can be divided into *passive* and *active* network sharing. Passive sharing refers to the reuse of components such as physical sites, tower masts, cabling, cabinets, power supply, air-conditioning, etc. Active sharing refers to the reuse of backhaul, base stations and antenna systems, the reuse of the latter two being labeled as active Radio Access Network (RAN) sharing. According to a market survey [3], mobile infrastructure sharing has been already deployed by over 65% of European operators in different ways, and this trend is expected to expand in the future.

Estimations regarding the expected savings for operators by implementing active network sharing have been conducted, see e.g., [4]. This study concluded that operators worldwide could reduce combined OPEX and CAPEX costs in up to \$60 Billion over a 5 years period through network sharing and at least 40% of these cost savings are expected to come from active network sharing.

The traditional model of single ownership of all network layers and elements is thus being challenged. While the most basic concepts related to Passive network sharing are easier to implement and have already been partially exploited, Active network sharing is raising in importance to enable substantial and sustainable reduction in network expenses and thus ensure operators' future competitiveness. The work presented in this paper addresses the aforementioned challenges by presenting a network model that possess desirable properties for RAN Sharing and presents both centralized and distributed solutions with analytical guarantees that efficiently allocate RAN resources fairly among operators.

II. Related work

Resource allocation optimization in wireless and cellular networks has been extensively studied. A common approach followed by many proposals is to provide load balancing by maximizing a certain objective function. The authors of [5] follow a max-min fairness criterion to ensure fairness and load balancing through association control. For multirate WLANs, [6] and [7] aim at providing proportional fairness to address the resource allocation problem, which is shown to be NP-hard, and propose centralized approximation algorithms that provide worst-case guarantee; in particular, [6] designs a 2+ϵ-approximation algorithm based on a reduction to the Generalized Assignment Problem (GAP). This algorithm relies on the 2-approximation for GAP proposed in [8], which has been overcome by [9] and recently further improved by [10].

In cellular networks, [11] formulates a proportional fairness problem specifically designed for 3G data networks and proposes an heuristic online algorithm to solve the optimal allocation in a network-wide context. In the same context, [12] designs and evaluates a distributed algorithm that converges to a near-optimal solution providing proportional fairness in an heterogeneous cellular networks. In [13], the authors propose and analyze an iterative distributed α-optimal algorithm for user association that converges to a globally optimal allocation. All the above focus on a network exploited by a single

The content of this document is part of a collaborative work. I would like to express thanks to Dr. Albert Banchs, Dr. Xavier Pérez-Costa, Dr. Gustavo de Veciana, Dr. Jorge Ortin and Dr. Konstantinos Samdanis for their valuable help on this work. A previous version of this work was accepted for oral presentation and publication in VTC-Spring 2015 [1].

operator, in contrast to our work which addresses Active RAN sharing between multiple operators.

Active RAN sharing enables pooling of radio resources enhancing the overall RAN utilization, while reducing infrastructure investments. 3GPP Services Working Group (WG) SA1 studied RAN sharing, analyzing a set of use cases deriving business requirements [14]. The architecture and operations that enable different MNOs to share the RAN are specified by the 3GPP Architecture WG SA2 in [2] detailing two different approaches. One refereed to as Multi-Operator Core Network (MOCN), where each operator is sharing eNBs connected to the core network elements of each MNO using a separate S1 interface. In the second one named Gateway Core Network (GWCN), operators share additionally the Mobility Management Entity (MME). GWCN enables extra cost savings compared to MOCN, but at the price of reduced flexibility, i.e. no mobility among different Radio Access Technologies and no interworking with legacy networks.

Recent efforts have been addressed to design RAN Sharing architectures based on Network Virtualization concepts [15]–[18]. Building on eNB virtualization, [15], [16] introduce the notion of Network Virtualization Substrate (NVS), a virtualization technique that provides an interface for operators. This interface allows to reserve wireless resources in slices, modifying the base stations MAC schedulers. The same authors enhanced NVS with the design of CellSlice [17], which adopts the same solution from a gateway-level perspective while enabling remotely controlled scheduling decisions, thereby simplifying near-term adoption. An SDN-based approach to control multi-tenant slices is proposed in [18] and a software defined radio access network simulator was designed by authors of [19]. All these works focus on architectural aspects of RAN Sharing and do not deal with the design of specific algorithms to share the resources, in contrast to the approach proposed in this paper which focuses on the algorithmic part. We adopt such capacity broker architecture and introduce intelligence in allocating resources and assigning user to particular base stations.

Like in this work, [20], [21] address the design of optimal algorithms for RAN sharing. In [21], the authors present RadioVisor, which aims at the optimization of the total network utility by using a max-min fairness criterion. RadioVisor, using simulator proposed in [19], proposes an heuristic algorithm to slice time-frequency slots at each base station and share these among N network operators while ensuring isolation. In contrast to RadioVisor, the approach that we propose in this paper relies on the proportional fairness criterion, which we have shown to provide many desired features. [20] proposes NetShare, a network-wide radio resource management framework that provides RAN Sharing. While [20] presents a proportional fairness formulation similar to the one we propose here, the authors do not provide a rationale to justify their choice, in contrast to the solid analytical arguments that we provide here. In order to solve the proportional fairness formulation, the authors use a general solver of non-linear problems, which incurs a very high computational complexity and can only be use for small scenarios; in contrast, we propose an efficient algorithm that provides some performance bound guarantees and can be used for large scenarios

To the best of our knowledge, this work is the first one that (i) provides an analytical proof for a set of desirable properties of the RAN Sharing resource allocation problem, achieving a trade off between optimality and fairness; and (ii) proposes an efficient algorithm that provides performance bound guarantees for this problem.

III. Network Model and Problem Formulation

In this section, we describe our assumptions, the utility function of an operator as well as the formulation of our optimization problem.

A. *Model and notation*

Our cellular LTE-Advanced scenario consists on multiple base stations where there are N users from K different operators sharing the network. For simplicity, we assume that traffic is elastic, and all users are saturated. MORA algorithm computes an optimal solution assuming that there is a centralized entity that has all information about the network. Table I presents all the notation symbols and terminology of our model.

Note that the algorithm proposed here could be combined with some other scheme that deals with inelastic traffic. Specifically, the technique for inelastic traffic would determine the resources received by each operator to satisfy its inelastic traffic needs, such us the one proposed in our previous work [1], and then the algorithm that we propose here would be used to share the remaining resources among elastic traffic.

TABLE I
NOTATION

Symbol	Definition
K	Number of operators
v_k	Share of operator k
U	Set of users
A	Set of base stations
N	Number of users: $N = \|U\|$
M	Number of base stations $M = \|A\|$
N_k	Number of users from operator k
N	Number of users from all operators
w_i	Weight of user i
r_i	Throughput of user i
b_{ij}	Fraction of resource blocks of BS j assigned to user i
c_{ij}	Transmission rate of user i in BS j
x_{ij}	Association of user i in BS j

B. *Sharing criterion for elastic traffic: Problem formulation*

In order to share the resources among users, we need to decide (i) how to associate users to base stations; and (ii) how to share the resources of each base station among its associated users. The main goal when doing this is to achieve both inter- and intra-operator fairness:

- **Inter-operator fairness**: the amount of resources assigned to each operator should be distributed proportionally the share of the operator v_k, which represents fraction of network resources to which the operator is entitled.

- **Intra-operator fairness**: the resources assigned to an operator should be fairly distributed among its users according to the operator's utility.

Following [22], which addresses a similar problem, we establish that the optimal configuration should be the one that maximizes the total welfare of the network, W: this function corresponds to the aggregation of the utilities of the different operators, where the utility of an operator corresponds in turn to the aggregation of the utility of its users, given by the throughputs of the users, r_i. Thus,

$$W = U_{inter}(U_{intra}(r_i))$$

where we refer to U_{inter} as the inter-operator utility and to U_{intra} as the the intra-operator utility or (simply) operator utility.

For the utility of an operator or intra-operator utility, we follow the *proportional fairness* criterion, which is a widely accepted criterion to provide fairness for elastic traffic,

$$U_k = U_{intra}(r_i) = \sum_{i=1}^{N_k} \log(r_i)$$

For the inter-operator utility, we sum the individual utilities of the different operators, weighted by some weight w_k that represents the priority that the users of a given operator are given in the resource allocation:

$$W = U_{inter}(U_k) = \sum_{k=1}^{K} w_i U_k = \sum_{k=1}^{K}\sum_{i=1}^{N_k} w_i \log(r_i)$$

In order to set the weights w_i, we follow the following rationale. According to Theorem 1, when two users of different operators compete for resources in the same base station, they share the base station's resources proportionally to their weights. Hence, the total amount of resources received by an operator will grow proportionally to the sum of the weights of all its users, $\sum_{i\in\mathcal{U}_k} w_i$. If we let the operator's share v_k represent the contribution of the operator to the overall network (like, e.g., the monetary contribution), then the total amount of resources received by the operator should be given by v_k. Thus, to ensure that resources are fairly shared among operators, we enforce that the following equation is satisfied:

$$\sum_{i\in\mathcal{U}_k} w_i = v_k$$

The above is interpreted as follows. The v_k represents the share of resources an operator is entitled to. This share is distributed among the operator's users, such that the weight of each user, w_i, represents the fraction of the operator's share assigned to the user. Thus, the sum of the w_i's has to add up to the share v_k. The operator may assign larger or smaller weights to the user depending on their relative priority. In the case the operator wants to give all users the same priority, we will have $w_i = v_k/N_k \ \forall i \in \mathcal{U}_k$.

With the above, we can formulate our optimization problem as follows:

$$maximize \qquad W = \sum_{k=1}^{K}\sum_{i=1}^{N_k} \frac{v_k}{N_k} \log(r_i) \tag{1a}$$

subject to:

$$r_i = \sum_{j=1}^{M} b_{ij} x_{ij} c_{ij} \tag{1b}$$

$$\sum_{j=1}^{M} x_{ij} = 1, \forall i \in N_k \tag{1c}$$

$$\sum_{k=1}^{K}\sum_{i=1}^{N_k} b_{ij} \cdot x_{ij} \le 1, \forall j \tag{1d}$$

$$x_{ij} \in \{0,1\} \tag{1e}$$

$$b_{ij} \ge 0 \tag{1f}$$

where the above equations impose the following constraints:

- (1b) gives the rate of a user as a function of the transmission rate and resources blocks assigned at the base station where it is connected; this is given by the base station to which the user is associated (x_{ij}), the fraction of resource blocks it receives from this base station (b_{ij}), and the transmission rate (c_{ij}), which is given by the modulation-coding scheme of the user at this base station.[1]
- (1c) forces to every user to be connected.
- (1d) forces that the fraction of resource blocks used in a base station cannot exceed 1 (i.e., the total capacity of the BS).
- (1e) imposes that the association variable $x_{i,j}$ is binary.
- (1f) imposes that the fraction of resource blocks b_{ij} is a positive real number.

We refer to the above strategy as the *multi-operator resource allocation* (MORA) problem. Next, we prove some properties for this strategy and evaluate the complexity of this problem.

IV. MORA PROPERTIES

In the following we show some desirable properties of the proposed resource allocation strategy, both in terms of overall performance, fairness and complexity. In terms of fairness, we would like to show that the proposed allocation is fair both in the sharing of resources of each base stations as well as in the way users are allocated to base stations.

A. BS Association

The following theorem shows that the resources of a base station are fairly shared among the operators using that base station. In particular, the fraction of resource blocks received by each user is proportional to its weight.

Theorem 1. *The fraction of resource blocks allocate to user i in base station j with the MORA strategy satisfies:*

$$b_{ij} = \frac{w_i x_{ij}}{\sum_{p=1}^{N} w_p x_{pj}}, \forall i, j \tag{2}$$

where w_i is the weight of user i – this weight depends on the operator: all the users of operator k have a weight equal to w_k, i.e., $w_i = w_k = v_k/N_k$.

[1] Note that c_{ij} is normalized to the total number of resource blocks of the base station, such that the throughput is given by the fraction of resource blocks multiplied by c_{ij}.

Proof: The theorem explains how the resources are allocated inside the base station and not the way users are allocated to them. Thus, supposing that we already have an association for x_{ij} of users x^*_{ij}, using Lagrangian methods, we can show that the above is satisfied. The lagrangian function for our problem is:

$$L(x,\lambda) = \sum_{i=1}^{N} w_i x^*_{ij} \log(r_i) + \lambda(1 - \sum_{i=1}^{N} b_{ij}) =$$
$$= \sum_{i=1}^{N} w_i x^*_{ij} \log(b_{ij}) + w_i x^*_{ij} \log(c_i) + \lambda(1 - \sum_{i=1}^{N} b_{ij})$$

Setting $\frac{\partial L}{\partial b_{ij}} = 0$ we get:

$$\frac{w_i x^*_{ij}}{b_{ij}} - \lambda = 0, \forall i$$

Then, since $\sum_{i=1}^{N} b_{ij} = 1, \forall j$, the unique optimal transmission resource allocation stands as follows:

$$b_{ij} = \frac{w_i x^*_{ij}}{\sum_{p=1}^{N} w_p x^*_{pj}}, \forall i, j$$

which proves the theorem. ■

It follows from the above theorem that the strategy followed by MORA allocates more resources to those operators whose users are at locations in which there is a lower demand from other users/operators. We refer to this notion, in which resources are allocated to operators based on the spatial distribution of their users, as *spatial fairness*. Another inherent property resulting from the above theorem relates the operator's share: as long as two operators use base stations with the same level of congestion (i.e., the same demand), the total amount of resources that they will receive will be proportional to their share. Lastly, the above theorem also indicates that only the fraction of users of operators and their shares have an influence on the amount of resources that receives from a base station and not the absolute number of users, so an operator that multiplies her users will not receive more resources unless increments her share (v_k). All this properties are formally proven in the corollaries of the appendix.

B. User association

So far we have looked at the fairness when sharing the resources of a base station among its users. Another desirable property from a fairness standpoint is on the user association.

The following theorem confirms the optimality in terms of overall performance of the resource allocation strategy proposed in the previous section.

Theorem 2. *The MORA allocation is Pareto-optimal.*

Proof: The proof follows easily from the fact that MORA chooses the configuration $\{b_{ij}, x_{ij}\}$ that maximizes $\sum_k w_k U_k$. This implies that for any alternative allocation $\{b'_{ij}, x'_{ij}\}$ for which $U'_l > U_l$ for some l, then necessarily we have $U'_m < U_m$. Indeed, if this was not the case than $\sum_k w_k U'_k$ would be larger than $\sum_k w_k U_k$, which would contradict our initial statement. This proves that the allocation resulting from MORA is Pareto-optimal. ■

In order to show that the user association provided is satisfactory for all operators (and hence fair), we would like the user association resulting from MORA represents a Nash equilibrium, meaning that an operator cannot obtain any gain by unilaterally deviating from the algorithm and following a different user association strategy. While we cannot prove that the proposed resource allocation is a Nash equilibrium, the following theorem shows that approximates a Nash equilibrium allocation.

Theorem 3. *If the resources blocks of all base stations are divided among their users according to (2), the user association resulting from MORA is an ϵ-approximate Nash equilibrium, with $\epsilon = v_k \log(e)$.*

Proof: Let us look at the maximum gain in utility that an operator can experience when one of his users, user j, is moved from one base station to another. The total gain in utility, which can not exceed 0, is given by the gain in utility by the operator, denoted by ΔU_{kj}, plus the difference in the utility of the other operators in the previous base station, p, and the new one, n:

$$\Delta U_{kj} + \sum_{i \in n} w_i \log\left(\frac{\sum_{i\in n} w_i}{w_k + \sum_{i \in n} w_i}\right)$$
$$+ \sum_{i\in p} w_i \log\left(\frac{w_k + \sum_{i\in n} w_i}{\sum_{i\in n} w_i}\right) \leq 0$$

where the above has been obtained taking into account that $log(r_i) = log(w_i) - log(\sum_j w_j) + log(c_{ij})$ and that the terms $log(w_i)$ and $log(c_{ij})$ are constant.

The above can be rewritten as

$$\Delta U_{kj} + w_k \log\left(\frac{\sum_{i\in n} w_i/w_k}{1 + \sum_{i\in n} w_i/w_k}\right)^{\sum_{i\in n} w_i/w_k}$$
$$+ w_k \log\left(\frac{1 + \sum_{i\in n} w_i/w_k}{\sum_{i\in n} w_i/w_k}\right)^{\sum_{i\in p} w_i/w_k} \leq 0$$

Given that $(x/(1+x))^x \leq 1$ and $((1+x)/x)^x \leq e$, for $x \geq 0$, we have

$$\Delta U_{kj} \leq w_k \log(e)$$

The above gives an upper bound on the increase of utility for a user. The increase of the total utility of the operator, ΔU_k, is upper bounded by the sum for all users,

$$\Delta U_k = N_k \Delta U_{kj} \leq N_k w_k \log(e) = v_k \log(e)$$

which proves the theorem. ■

C. Complexity & Uniqueness

The MORA problem is a *non-linear integer programming problem*. The following theorem shows that this problem is NP-hard and hence there is no algorithm able to solve the problem in polynomial time unless $P = NP$.

Theorem 4. *The MORA problem is NP-hard.*

Proof: The reduction is via the 3-dimensional matching problem which is known to be NP-complete. Recall that the 3-dimensional matching problem is stated as follows. Let us consider disjoint sets $C = \{c_1, \ldots, c_n\}$, $D = \{d_1, \ldots, d_n\}$ and $E = \{e_1, \ldots, e_n\}$, and a family $F = \{F_1, \ldots, F_m\}$ of triples with $|F_i \cap C| = |F_i \cap D| = |F_i \cap E| = 1$ for $i = 1, \ldots, m$, with $m \geq n$. The question is whether F contains a matching, i.e., a subfamily F' for which $|F'| = n$ and $\cup_{F_i \in F'} F_i = C \cup D \cup E$.

Our reduction is along the lines of [11]. We call the triples that contain c_j *triples of type j*. Let f_j be the number of triples of type j for $j = 1, \ldots, n$. BS i corresponds to the triples F_i for $i = 1, \ldots, m$. We create two types of users, element users and dummy users. We have $2n$ element users, corresponding to the $2n$ elements of $D \cup E$. There are $f_j - 1$ dummy users of type j for $j = 1, \ldots, n$. Note that the total number of dummy users is $m - n$. Element users can connect to the BSs that correspond to a triple that contains this element, with an average rate of R. Dummy users of type j can connect (also with an average rate of R) to the BSs that correspond to triples of type j. Element users have a weight $w = 1$ and dummy users have a weight $w = 2$. We claim that a matching exists if and only if the MORA's optimal objective function is $W = 2n \log(R/2) + 2(m - n) \log(R)$.

The value of the objective function is bounded above by the following optimization problem:

$$maximize \qquad \sum_{i=1}^{2n} \log(b_i R) + \sum_{i=1}^{m-n} 2 \log(b_i R) \qquad (3)$$

subject to $\sum_{i=1}^{2n} b_i + \sum_{i=1}^{m-n} b_i = m$, where the first term of the summation corresponds to the element users and the second term to the dummy users, and b_i is the fraction of resource blocks assigned to a user.

By applying the Lagrange multiplier method, it can be easily seen that the above optimization problem is solved when $b_i = 1/2$ for the element users and $b_i = 1$ for the dummy users. This gives an upper bound on W equal to $2n \log(R/2) + 2(m - n) \log(R)$. This corresponds to a global maximum, and thus any other set of b_i values yields a smaller W.

Assume that there is a matching. For each $T_i = (c_j, d_k, e_l)$ in the matching, we associate element users d_k and e_l with BS i. For each j, this leaves $f_j - 1$ idle BSs corresponding to tripes of type j that are not in the matching. We associate the $f_j - 1$ dummy users of type j to these $t_j - 1$ BSs. This assignment has an objective function of $W = 2n \log(R/2) + 2(m - n) \log(R)$, which is equal to the upper bound given above. In case there is no matching, it is not possible to have the $2n$ element users sharing n BSs with $b_i = 1/2$ each, and therefore we cannot achieve the distribution of b_i values that maximizes W. According to the above result, this implies that we obtain a smaller W value, which proves the theorem. ■

Other interesting characteristic is that the problem can have multiple optimal solutions, therefore it's possible to prove the non-uniqueness property of the problem, which is formally expressed in the following theorem.

Theorem 5. ***(Non-Uniqueness).*** *MORA problem can have multiple optimal solutions.*

Proof: The proof follows directly analyzing an instance of the problem where a user u can move from a base station, i.e. n, to another base station p without having an impact on the network utility K, or what is the same, causing that $\Delta W = 0$. Let ΔU_j be the sum of the utilities for every operator in base station j ($\Delta U_j = \sum_{k \in K} \Delta U_{kj}$), then:

$$\Delta W = \sum_{j \in M} \Delta U_j = \Delta U_n + \Delta U_p$$

Equivalent to the proof for Theorem 3 we can present this differences taking into account that $\log(r_i) = \log(w_i) - \log(\sum_j w_j) + \log(c_{ij})$ and that the terms $\log(w_i)$ and $\log(c_{ij})$ are constant, and also supposing that the user can connect to both stations with the same rate, i.e. ($c_{un} = c_{up}$), which leads into:

$$\begin{aligned}\Delta W = \Delta U_n + \Delta U_p = \sum_{i \in n} w_i \log\left(\frac{\sum_{i \in n} w_u}{w_u + \sum_{i \in n} w_i}\right) \\ + \sum_{i \in p} w_i \log\left(\frac{w_u + \sum_{i \in p} w_i}{\sum_{i \in p} w_i}\right) = \\ = \sum_{i \in n} w_i \left[\log\left(\sum_{i \in n} w_i\right) - \log\left(w_u + \sum_{i \in n} w_i\right)\right] \\ + \sum_{i \in p} w_i \left[-\log\left(\sum_{i \in p} w_i\right) + \log\left(w_u + \sum_{i \in p} w_i\right)\right]\end{aligned}$$

In the case when $\sum_{i \in n} w_i = \sum_{i \in p} w_i$, since w_u is the same for all base stations, $\Delta W = 0$ and there are two possible associations of user u that gives the same utility, proving the theorem. ■

V. Multi-Operator Resource Allocation (MORA) Centralized Algorithm

Due to the NP-Hardness of the MORA problem, we propose a polynomial time algorithm similar to the one proposed in [6], which solves this problem while ensuring performance guarantees with an additive ratio. The proposed algorithm is based on a **discretized relaxation and reduction to Generalized Assignment Problem (GAP)**, a well know problem. An schematic view of the algorithm is given in Figure 1, which included a description of the algorithm steps.

1. Discretize scheduling period solving relaxation.
   $y = solve\_relaxDLP(\{c_{ij}\}, D)$
2. Transform into a GAP using (x, b) and u.
3. Solve problem by a rounding technique
   $\hat{x} = roundGAP(b, x, u)$
4. Redistribute resource blocks percentages

Fig. 1. Approximation algorithm

A. Discretization

In our original MORA problem, we are considering that the percentage of resources blocks assigned to a user is a continuous value between 0 and 1, assumption that doesn't stand in reality since the amount of resource blocks is an integer number ($b \in \mathbb{Z}$). Considering this real world constraint, and in order to reduce the complexity of our problem, we discretize the scheduling period, moving from a continuous possible resource blocks distribution to a discrete one and by that, linearizing the objective function of our problem. We introduce a new indicator variable $y_{ij\tau}$, which is equal to 1 if and only if user i is assigned to access point j, and access point j has allocated τ (out of D) percent of the resource blocks to user i. The value of τ ranges from 1 to D. With this new variable, the formulation of the discretized problem is defined as follows:

$$maximize \quad \sum_{k=1}^{K}\sum_{i=1}^{N_k}\sum_{j=1}^{M}\sum_{\tau=1}^{D}\frac{v_k}{N_k}y_{ij\tau}\log\left(\frac{\tau}{D}\cdot c_{i,j}\right) \tag{4a}$$

subject to:

$$\sum_{j=1}^{M}\sum_{\tau=1}^{D} y_{ij\tau} = 1, \forall i \in N \tag{4b}$$

$$\sum_{k=1}^{K}\sum_{i=1}^{N_k}\sum_{\tau=1}^{D}\frac{\tau}{D}y_{ij\tau} = 1, \forall j \tag{4c}$$

$$y_{ij\tau} \in \{0,1\} \tag{4d}$$

Let $W_{MORA}(x,b)$ stand for the solution of the original (non-linear, NP-Hard) MORA problem, and $W_{Dis}(y)$ the solution to the above (discretized) algorithm. Then:

Lemma 1. *If we choose a big enough D:*

$$W_{Dis}(y) \geq W_{MORA}(x, b/(1+\delta))$$

Proof: If the resolution of the discretized version of the problem allows to represent every single case of the non-dicretized problem, the loss introduced by the discretization is 0. To achieve this is neccesary to set a big enough amount of steps for the time scheduling that can be obtained by considering the worst association case.

$$D = (1+\delta)\cdot\frac{\sum_{k=1}^{K}\sum_{i=1}^{N_k} v_k}{min(v_k)} \tag{5}$$

∎

Although we transformed the objective function into a non-linear function, the problem is combinatorial due to the binary variable $y_{ij\tau}$, and hence, the computation is essentially impossible for even a modest-sized cellular network. To overcome this, we can relax the user association by multiple-BS user association to compute an integer relaxation of the linear problem allowing $\hat{y}_{ij\tau}$ take fractional values. This physical relaxation gives an upper bound solution for the problem and it's then direct that any solution of $W_{Dis}(y)$ is feasible for the fractional association problem $W_{Dis}(\hat{y})$ leading into $W_{Dis}(\hat{y}) \geq W_{Dis}(y)$. It's as well direct that a fractional solution (x'', b'') for the original MORA formulation follows $W_{MORA}(x'', b'') \geq W_{Dis}(\hat{y})$

Lemma 2. *Joining previous statements we can ensure that:*

$$W_{MORA}(x'', b'') \geq W_{Dis}(\hat{y}) \geq W_{MORA}(x, b/(1+\delta))$$

B. Reduction to Generalized Assignment Problem

Once we have obtained a solution for the relaxed discretized version of the problem, we proceed to reduce our problem to a GAP, by fixing the resource block assignation and utility and only standing x_{ij} as decision variables. Thus, the reduction to GAP it can be easily done by holding resource block and utility by using result obtained from previous step, $\hat{y}$.

Set: $\forall i \in U, j \in A$:

$$x_{ij} = \sum_{\tau=1}^{D}\hat{y}_{ij\tau} \tag{6}$$

$$p_{ij} = \frac{\sum_{\tau=1}^{D}\frac{\tau}{D}\hat{y}_{ij\tau}}{x_{ij}} \tag{7}$$

$$u_{ij} = \frac{v_k}{N_k}\log(c_{ij}b_{ij}) \tag{8}$$

C. Solve problem by rounding technique

Once this transformation is done, our problem fits with the standard formulation of GAP. It is well known that GAP is NP-Hard [23], but we can use a rounding algorithm to transform our multiple-BS association fractional solution to an single-bs association integer solution. To do this, we use a rounding scheme proposed by [8] in order obtain the single bs-association solution. This rounding technique is a 2-approximation algorithm for the maximization GAP that takes a fractional assignment $\hat{x}$ as input, and constructs an integer assignment vector x with three properties: (i) the objective function value for the integer association is greater than the one obtained with the fractional association, (ii) each user i is assigned to one machine and (iii) the load imposed on each server is at most 2 ($\sum_{i=1}^{N_k} b_{i,j}\cdot x_{i,j} \leq 2$) for every base station.

D. Redistribute resource blocks percentage

Once we have the single-bs association solution for our problem what left is reassign time assigned to each of the users $\hat{b}$ in the way of equation (2). This reassignment give us a solution with the properties described below.

1) The vector x in a feasible (fractional) solution (x'', b'') to the formulation MORA is feasible for the formulation GAP. Alternatively, the feasible solution x for GAP together with its vector $\hat{b}$ is a feasible solution to MORA:

$$W_{GAP}(\hat{x}) = W_{MORA}(x'', b'') \geq W_{Dis}(\hat{y}) \geq W_{MORA}(x, b/(1+\delta))$$

2) The solution $(\hat{x}, \{\hat{t}\})$ may not be feasible, as some of the access points may be over-scheduled by a factor of 2. However, if we scale down the time allocations $\hat{t}' = \frac{1}{2}\hat{b}$, we obtain a feasible solution with the following property:

$$W_{GAP}(\hat{x}, \{\hat{b'}\}) \geq W_{MORA}(x, \frac{b}{2\cdot(1+\delta)})$$

Theorem 6. *With the proposed centralized algorithm, the total welfare of the network satisfies $W \geq W^*_{MORA}\sum_{i=1}^{N} w_i \log(2)$, where W^*_{MORA} is the total welfare of the optimal allocation.*

Proof: Consider any integral association $W_{MORA}(x^*, b^*)$ that is a feasible solution to 1. Then, for the solution $W(\hat{x}, \hat{b})$ produced by our algorithm it holds that:

$$W(\hat{x}, \hat{b}) \geq W_{MORA}(x^*, \frac{b^*}{2+O(\delta)}) \quad (9)$$

We can transform this inequality into an additive approximation ratio as follows:

$$\begin{aligned} W(\hat{x}, \hat{b}) &= \sum_{k=1}^{K}\sum_{i=1}^{N_k}\sum_{j=1}^{M} \hat{x_{ij}} \cdot w_i \cdot \log(c_{ij}\hat{b_{ij}}) \geq \\ &\geq \sum_{k=1}^{K}\sum_{i=1}^{N_k}\sum_{j=1}^{M} x^*_{ij} \cdot w_i \cdot \log(c_{ij}\frac{b^*_{ij}}{2}) = \\ &= \sum_{k=1}^{K}\sum_{i=1}^{N_k}\sum_{j=1}^{M} x^*_{ij} \cdot w_i \cdot \log(c_{ij}b^*_{ij}) - \\ &\quad - \sum_{k=1}^{K}\sum_{i=1}^{N_k}\sum_{j=1}^{M} x^*_{ij} \cdot w_i \cdot \log(2) = \\ &= W_{MORA}(x^*, b^*) - \\ &\quad - \sum_{k=1}^{K}\sum_{i=1}^{N_k}\sum_{j=1}^{M} x^*_{ij} \cdot w_i \cdot \log(2) = \\ &= W_{MORA}(x^*, b^*) - \sum_{i=1}^{N} w_i \log(2) \end{aligned}$$

Last equality holds since $\sum_{i=1}^{N_k}\sum_{j=1}^{M} x_{ij} = 1$. Then is proven that:

$$W(\hat{x}, \hat{t}) \geq W_{MORA}(x^*, b^*) - \sum_{i=1}^{N} w_i \log(2) \quad (10)$$

■

VI. Multi-Operator Resource Allocation (MORA) distributed algorithm

In LTE-A there is no interface to gather all the data of the network at a centralized location in a timely manner. There are interfaces to gather information from neighbours. Therefore, an interesting option is to implement an algorithm that optimizes performance in real-time is by means of a distributed algorithm.

In the following, we design an algorithm to maximize the total welfare that satisfies $W \geq W^* - \sum_{i=1}^{N} w_i \log(e)$, where W^* is the total welfare with the optimal strategy. It is worthwhile noting that our centralized algorithm provides a slightly tighter bound, namely $W \geq W^* - \sum_{i=1}^{N} w_i \log(2+\epsilon)$. However, as mentioned above, that algorithm is centralized and much more complex to use in real world scenarios.

The proposed algorithm is based on a greedy approach, in which in each iteration a given user i reassociates to the base station that provides her with the largest r_i. The following theorem shows that this algorithm satisfies the finite improvement property, which states that an arbitrary sequence of the algorithm executed by each user always converges to an equilibrium in a finite number of steps.

Theorem 7. *The proposed distributed algorithm has the finite improvement property.*

Proof: The throughput of user i, r_i, is an increasing function of $c_{ij}/\sum_{l=1}^{N} w_j x_{jl}$. Hence, our algorithm can be understood as a congestion game in which the load at a base station is given by the sum of weights of the users at the base station, $l_j = \sum_j w_j$, and a user seeks to minimize $a_{ij} l_j$, where $a_{ij} = 1/c_{ij}$. This congestion game falls in the category of a singleton weighted congestion game with separable preferences, resource-independent weights (the w_i values are the same for all base stations) and linear variable cost (as l_j is linear with the w_i's). According to Theorem 4 of [24], such a game has the finite improvement property. ■

The following theorem shows that the solution at which the distributed algorithm converges provides a performance that is above a given lower bound.

Theorem 8. *With the proposed distributed algorithm, the total welfare of the network satisfies $W \geq W^* - \sum_{i=1}^{N} w_i \log(e)$, where W^* is the total welfare of the optimal allocation.*

Proof: Let j be the base station to which user i is associated with the distributed algorith, and k the base station to which it is associated with the optimal strategy. Since with the distributed algorithm we maximize r_i, the following holds:

$$w_i \log\left(\frac{w_i c_{ij}}{\sum_{j\in\mathcal{J}_d} w_j}\right) \geq w_i \log\left(\frac{w_i c_{ik}}{\sum_{k\in\mathcal{K}_d} w_k + w_i}\right)$$

where $\mathcal{J}_d$ and $\mathcal{K}_d$ are the set of users associated to base stations j and k, respectively, with the distributed algorithm.

Let $\mathcal{K}_{opt}$ denote the set of users associated to base station k with the optimal strategy. Since we stated above that $w_i \in \mathcal{K}_{opt}$, the following holds:

$$w_i \log\left(\frac{w_i c_{ij}}{\sum_{j\in\mathcal{J}_d} w_j}\right) \geq w_i \log\left(\frac{w_i c_{ik}}{\sum_{k\in\mathcal{K}_d} w_k + \sum_{k\in\mathcal{K}_{opt}} w_k}\right)$$

If we let r_i^* and r_i denote the throughput of user i with the optimal strategy and with the distributed algorithm, respectively, from the above it follows that

$$\begin{aligned} w_i \log(r_i^*) - w_i \log(r_i) \leq w_i \log\left(\frac{w_i c_{ik}}{\sum_{k\in\mathcal{K}_d} w_k}\right) - \\ w_i \log\left(\frac{w_i c_{ik}}{\sum_{k\in\mathcal{K}_d} w_k + \sum_{k\in\mathcal{K}_{opt}} w_k}\right) = \\ -w_i \log\left(\frac{\sum_{k\in\mathcal{K}_{opt}} w_k}{\sum_{k\in\mathcal{K}_d} w_k + \sum_{k\in\mathcal{K}_{opt}} w_k}\right) \end{aligned}$$

Summing the above over all users yields

$$W^* - W \leq -\sum_{i=1}^{N} w_i \log\left(\frac{\sum_{k\in\mathcal{K}^i_{opt}} w_k}{\sum_{k\in\mathcal{K}^i_d} w_k + \sum_{k\in\mathcal{K}^i_{opt}} w_k}\right)$$

where $\mathcal{K}^i_{opt}$ and $\mathcal{K}^i_d$ are the following set of users. Let us consider the base station to which user i is associated with the optimal strategy. Then, $\mathcal{K}^i_{opt}$ is the set of users associated to this base station with the optimal strategy, and $\mathcal{K}^i_d$ is the set of users associated to this base station with the distributed algorithm.

By rearranging terms, we can rewrite the above as

$$W^{*}-W \leq -\sum_{j=1}^{M}\sum_{k\in\mathcal{K}^{j}_{opt}} w_k \log\left(\frac{\sum_{k\in\mathcal{K}^{j}_{opt}} w_k}{\sum_{k\in\mathcal{K}^{j}_{d}} w_k+\sum_{k\in\mathcal{K}^{j}_{opt}} w_k}\right)$$

where $\mathcal{K}^{j}_{opt}$ is the set of users at base station j with the optimal strategy, and $\mathcal{K}^{j}_{d}$ is the set of users at base station j with the distributed algorithm.
From the above:

$$W^{*}-W \leq -\sum_{j=1}^{M}\log\left(\frac{\sum_{k\in\mathcal{K}^{j}_{opt}} w_k}{\sum_{k\in\mathcal{K}^{j}_{d}} w_k+\sum_{k\in\mathcal{K}^{j}_{opt}} w_k}\right)^{\sum_{k\in\mathcal{K}^{j}_{opt}} w_k}$$

$$= -\sum_{j=1}^{M}\sum_{k\in\mathcal{K}^{j}_{d}} w_k \log\left(\frac{\sum_{k\in\mathcal{K}^{j}_{opt}} w_k/\sum_{k\in\mathcal{K}^{j}_{d}} w_k}{1+\sum_{k\in\mathcal{K}^{j}_{opt}} w_k/\sum_{k\in\mathcal{K}^{j}_{d}} w_k}\right)^{\frac{\sum_{k\in\mathcal{K}^{j}_{opt}} w_k}{\sum_{k\in\mathcal{K}^{j}_{d}} w_k}}$$

Given that $(x/(1+x))^{x} \geq 1/e$ for $x \geq 0$, we obtain the following bound

$$W^{*}-W \leq \sum_{i=1}^{N} w_i \log(e)$$

which proves the theorem. ■

VII. Performance Evaluation

In order to evaluate the different algorithms proposed, in the following we describe the simulation setup used for performance benchmarking.

A. Simulation Setup

Mobile Network Model: The mobile network scenario considered, based on the IMT-Advanced evaluation guidelines [25], is depicted in Figure 2. It consists of 19 base stations in a hexagonal cell layout with 3 sector antennas. We focus on the 'Urban Micro-cell scenario' to model a dense 'small cell' deployment. The detailed system configuration parameters are summarized in Table II.

Mobility model: Users mobility follows the SLAW model defined in [26], which is a human walk mobility model based on real GPS-tracking measurements of more than 6 million UEs. This model allows us to synthetically generate UEs movements in our scenario in a realistic way. The configuration parameters used to generate the UEs movement in our experiments are given in Table III, and the resulting distribution depicted in Figure 2.

Signal power and c_{ij} estimation: In order to compute c_{ij}, we need to estimate the Signal Interference to Noise Ratio for each user i to each BS j following the idea proposed by [12], c_{ij} is computed as:

$$c_{ij} = f(\text{SINR}_{ij})$$

$$\text{SINR}_{ij} = \frac{P_j g_{ij}}{\sum_{k\in A, k\neq j} P_j g_{ij} + \sigma^2}$$

where P_j is the transmit power of BS j, g_{ij} denotes the channel gain between user i and BS j, which includes path loss, shadowing and antenna gain, and σ^2 denotes the thermal

Parameters	Values
Mobile Network	1 Tier, 19 BSs with 3 Sectors
Virtual Operators	3
IMT-A Scenario	Urban Micro (UMi)
Intersite distance	200 m
Users Mobility	SLAW Model
Carrier Frequency (f_c)	2.5 GHz
Simulation Bandwidth (BW)	10 MHz
Thermal noise level (σ^2)	-174 dBm/Hz
Pathloss	$36.7\log_{10}(d)+22.7+26\log_{10}(f_c)$
Antenna tx power (P_j)	41 dBm
Antenna gain	17 dBi
Shadow fading std	4 dB
Simulation Time	1 hour
Warm-up Time	10 minutes

TABLE II
ITU IMT-Adv Simulation Parameters Configuration

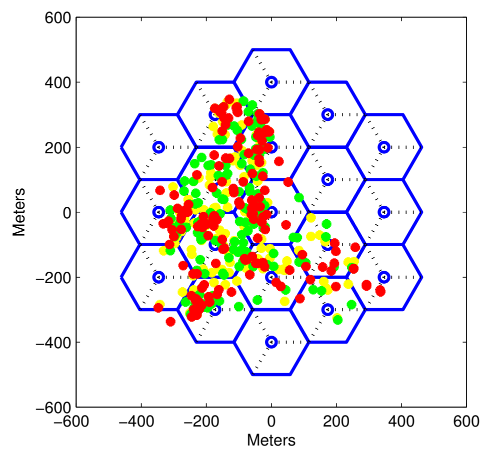

Fig. 2. User distribution models for 550 users. The different colors represent users from three different operators

noise power level. We use a shadowing factor represented by a log-normal function of mean 0 and standard deviation of 4 dB. Thus, under the assumption that the user achieves the Shannon capacity rate, the value of the rate stands:

$$\begin{aligned} c_{ij} &= f(\text{SINR}_{ij}) = log_2(1+\text{SINR}_{ij}) \\ &= log_2\left(1+\frac{P_j g_{ij}}{\sum_{k\in A,k\neq j} P_j g_{ij}+\sigma^2})\right) \end{aligned} \quad (11)$$

This value c_{ij} represent the rate in bps/Hz achievable by user i in base station j. If we want to compute the throughput in Mbps we need to make use of the bandwidth BW:

$$Throughput_{ij} = BW \cdot c_{ij}$$

B. Benchmark Algorithms

Benchmark 1 - Distributed Static Legacy Algorithm : The first algorithm considered in our study for benchmarking purposes is the *Distributed Static Legacy*. With this approach, UEs from different mobile network virtual operators are associated to the base station with the highest SINR. This models legacy systems with no multi-tenant support for UE re-association where the enforcement of the resource allocation per tenant would be performed at every base station independently at scheduling level (see [16] for an example).

Parameter	
Distance alpha	5
Waypoints(wp)	500
Inverse self-similarity of wp	0.95
Time (hours)	3
Clustering range (m)	50

TABLE III
SLAW PARAMETERS CONFIGURATION

Benchmark 2 - MORA and Upper bound of MORA : The second algorithm considered in our study for benchmarking purposes is *MORA*. It was previously discussed that this algorithm is NP-Hard and posses high computational cost. For this reasons all the results for this benchmark algorithm for a number of users greater than 130 came from an upper bound computed by a discretization of the logarithm as done in our centralized algorithm.

C. Performance Results

We next present the performance results obtained with the above simulation setup. We start by analyzing the performance from the perspective of the mobile infrastructure provider (e.g., operator with full management access to the physical infrastructure), focusing on the overall network performance, evaluating the utility for the different algorithms as well as the file transfer delay and throughput improvements due to our algorithm. Then, we present the results from the mobile (virtual) operators perspective side, looking at the performance of each operator and showing that proposed algorithms meet the network sharing agreements. Lastly, we conclude the theoretical complexity discussion adding some simulation results confirming the theoretical results.

Utility performance: For this experiment, we present the gains in terms of utility that our algorithms provides with respect to the current Legacy system (Static Slicing). In figure 3, we can see that our algorithms both centralized and distributed outperforms the utility achieved by Static slicing for different number of users in the system. It is also noticeable that both proposed algorithms have a performance close to the optimal (displayed in green), achieving the centralized the closer performance. The error bars in this figure represent the 95% confidence intervals for the utility in this scenario. [2]

Download time: One of the benefits introduced by our model is the one related with intra-operator fairness. Proportional fairness permit to users with low rates to receive a higher number of resources in order to exponentially improve the QoS experienced by them. In the figure 4, we have displayed the throughput for each of the 50 users of the system when we used Static Slicing (left) and our MORA centralized solution (right).

We can observe that the throughput values are more balanced in the case of our centralized solution (right), where the users with high rates transfer resource blocks to the users

[2]Error bars are not displayed in further figures for readability reasons.

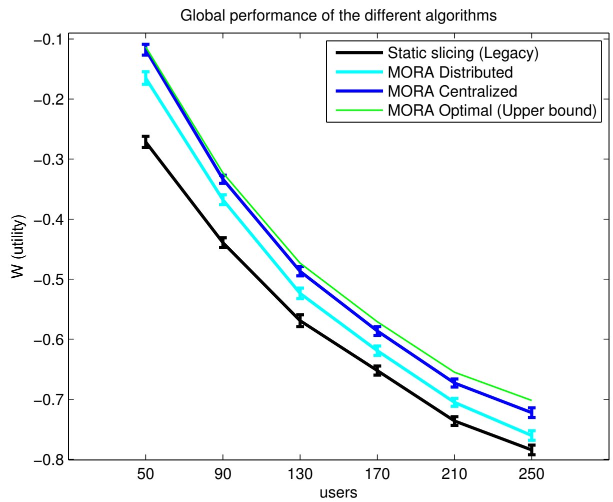

Fig. 3. Algorithm comparison of utility for different number of users in the system

with lower rates as is the objective of the proportional fairness paradigm.

As we pointed out, this have a clear relevance in the performance experienced by the users. In the figure 5, we have displayed the file transfer delay for a 20 megabytes archive by each of the 50 users of the system when we used Static Slicing (left) and our MORA centralized solution (right). Again a more uniform distributed file transfer delay is earned when MORA centralized algorithm is used in the user association. E.g. while in the case of static slicing 3 users out of the 50 (labeled 8, 18, 22) consume one minute or more, using the MORA algorithm none of them requires more than 60 seconds to download the file.

Shares agreements: From the operator perspective, the critical aspect regarding RAN sharing relies on share agreement fairness. Each operator buys a percentage of the total network share, defined in our paper as w_k, and the operator would like to receive this percentage of resources from the network. Our MORA solution respects the shares of the operators, as was discussed theoretically, and this is reflected in figure 6. In this figure, a comparison between resources received with the association $\sum_{i \in k} b_{ij}$ (blue) and percentage of network contracted v_k (green) is displayed for an scenario with three operators. Three different agreement cases were considered.

1) Uniform share agreement among the 3 operators $v_k = (1/3, 1/3, 1/3)$. (left subfigure)
2) Triangular share agreement among the 3 operators $v_k = (1/5, 3/5, 1/5)$. (center subfigure)
3) Irregular share agreement among the 3 operators $v_k = (1/10, 6/10, 3/10)$. (right subfigure)

In all the three figures, we can see that our formulation respects the agreements among the operators. It's worthwhile to note that the small deviations are due to the distribution of the users and for a uniform distribution $\sum_{i \in k} b_{ij} = v_k$.

Computational complexity: Lastly, the computational complexity of the proposed and benchmark algorithms was eval-

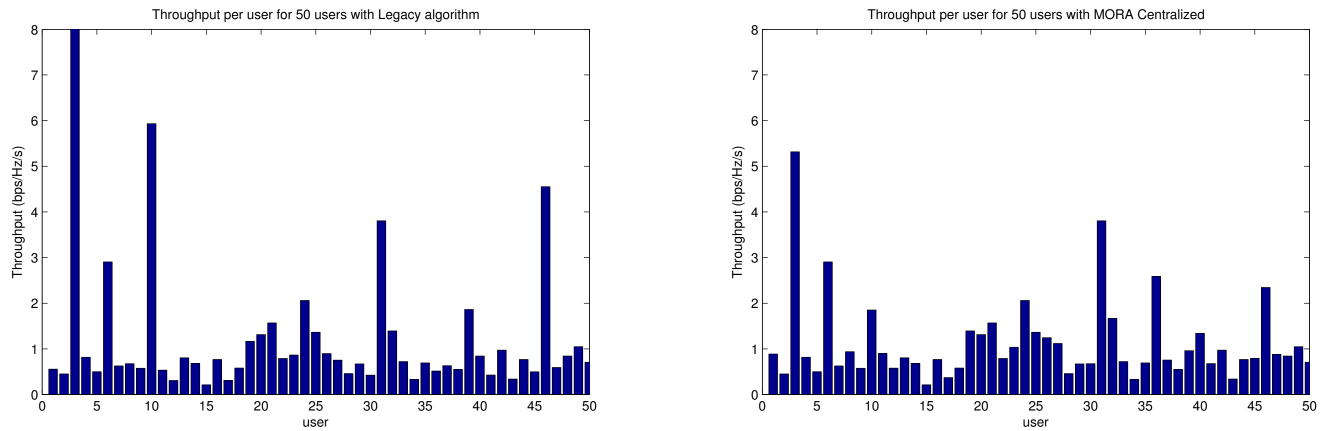

Fig. 4. Comparison of different algorithms throughput per user for 50 users.

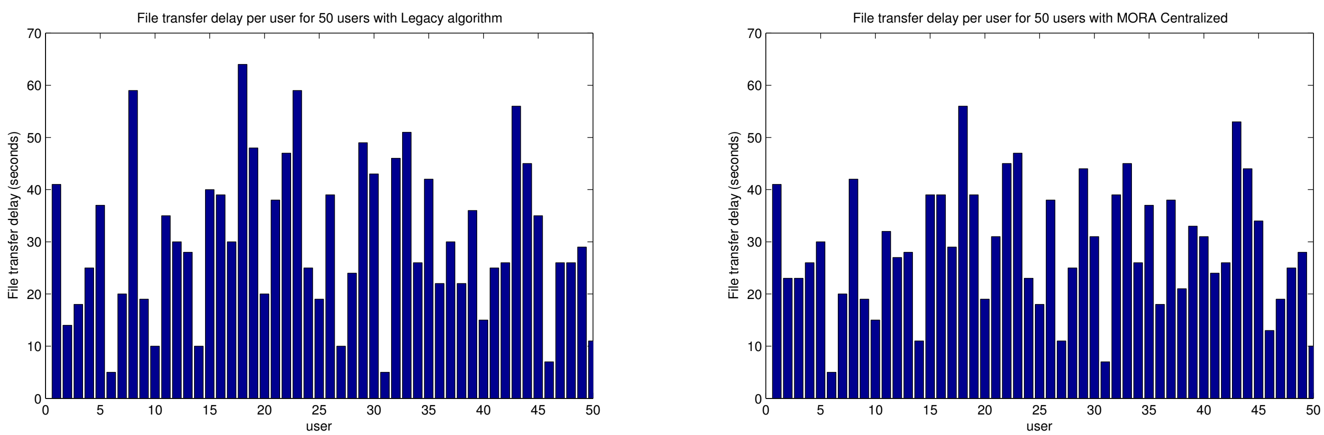

Fig. 5. Comparison of different algorithms file transfer delay per user for 50 users.

uated for different problem sizes and this comparison is depicted in figure 7. For the case of MORA optimal the consumed time is computed for the discretized version and not for the non-linear version, which is even more complex. In this figure we can see that MORA optimal has an exponential complexity growth due to her integer variables nature. The rest of the algorithms run in polynomial time and it's clear to see that their complexity is polynomial being the MORA centralized algorithm the more complex followed by the MORA distributed. The algorithm with less complexity cost is the highest SINR association Static Slicing, as expected. This confirms that both solutions proposed are computationally feasible thanks to the polynomial time solvable character.

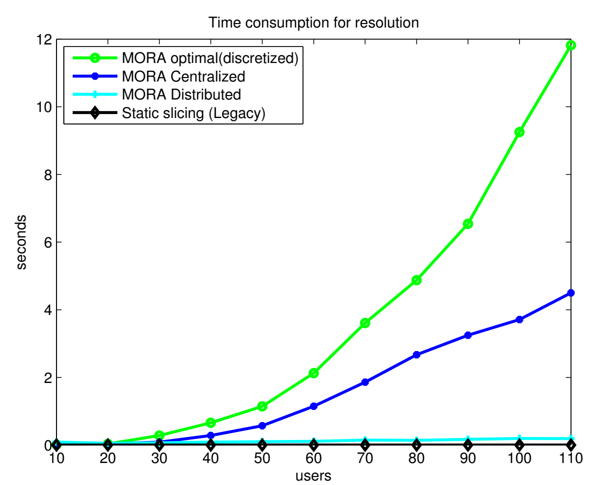

Fig. 7. Algorithm comparison of convergence time consumption for different problem sizes

VIII. Conclusions and Future Work

Multiple Operator Resource Allocation *MORA* network active sharing is turning into a key business and a fair approach for this assignment among different tenants needs a consistent framework. We presented an offline solution that copes with most of the desirable properties that this problems requires. Our solution takes into account both inter and intra operator fairness. This give to our solution a set desirable properties from base station association and user association perspectives respecting the percentages of the networks contracted by each operator and being proportionally fair with the users. We proved that *MORA* is an NP-Hardness problem and accordingly we propose a centralized approximation algorithm that provides performance guarantees with an additive ratio of $\sum_{i=1}^{N} w_i \log(2)$ and runs in polynomial time. In order to be able to optimize performance in real-time, we make use of the real-time neighbour information provided

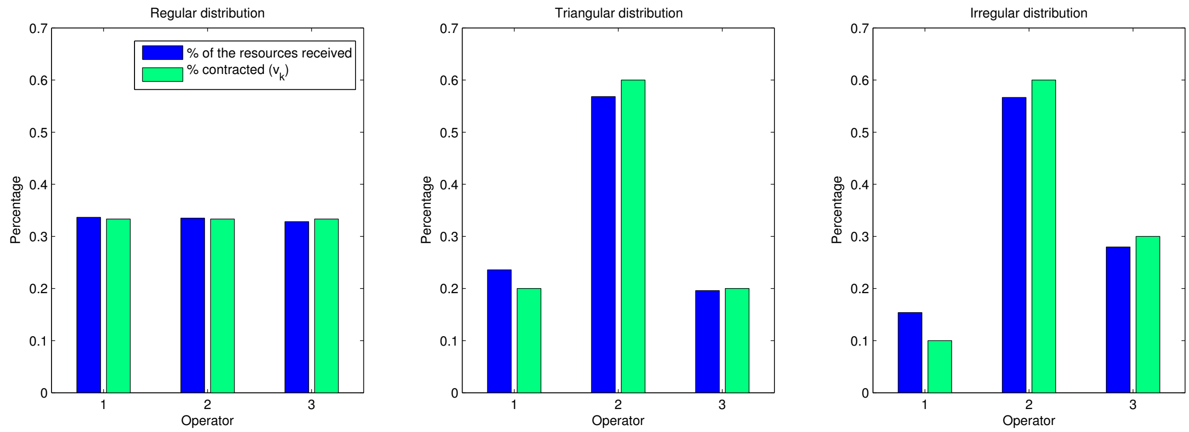

Fig. 6. Shares vs resources received for three different agreement cases.

by the LTE-Advanced interfaces (e.g. X1 interface) to design a distributed algorithm with a theoretical additive ratio of $\sum_{i=1}^{N} w_i \log(e)$. Finally, all the theoretical results detailed above has been proven by a set of performance evaluations in a LTE-Advanced system-level simulator. Then, this work is placing the first stone for efficient Active RAN Sharing. Even though there are a set of open issues for this problem such as the design of an offline optimal algorithm, an analysis of the robustness of our solution against any possible operator cheating decision or a real implementation of the solution proposed.

REFERENCES

[1] P. Caballero Garcés, X. Costa Pérez, K. Samdanis, and A. Banchs, "RMSC: A cell slicing controller for multi-tenant mobile networks," in *Vehicular Technology Conference (VTC 2015-Spring), 2015 IEEE*. IEEE, 2015.

[2] "Network Sharing; Architecture and Functional Description." 3GPP TS 23.251, v12.1.0, Jun. 2014.

[3] "Mobile Network Sharing Report 2010-2015, Development, Analysis & Forecasts, Market Study," Visiongain, 2010.

[4] "Active RAN Sharing Could Save $60 Billion for Operators." http://www.cellular-news.com/story/36831.php., 2009.

[5] Y. Bejerano, S. jae Han, and L. Li, "Fairness and load balancing in wireless lans using association control," *Networking, IEEE/ACM Transactions on*, vol. 15, no. 3, pp. 560–573, June 2007.

[6] L. Li, M. Pal, and Y. R. Yang, "Proportional fairness in multi-rate wireless lans," in *INFOCOM 2008. The 27th Conference on Computer Communications. IEEE*. IEEE, 2008.

[7] W. Li, S. Wang, Y. Cui, X. Cheng, R. Xin, M. Al-Rodhaan, and A. Al-Dhelaan, "Ap association for proportional fairness in multirate wlans," *Networking, IEEE/ACM Transactions on*, vol. 22, no. 1, pp. 191–202, Feb 2014.

[8] D. B. Shmoys and É. Tardos, "An approximation algorithm for the generalized assignment problem," *Mathematical Programming*, vol. 62, no. 1-3, pp. 461–474, 1993.

[9] L. Fleischer, M. X. Goemans, V. S. Mirrokni, and M. Sviridenko, "Tight approximation algorithms for maximum general assignment problems," in *Proceedings of the seventeenth annual ACM-SIAM symposium on Discrete algorithm*. ACM, 2006, pp. 611–620.

[10] U. Feige and J. Vondrak, "Approximation algorithms for allocation problems: Improving the factor of 1-1/e," in *Foundations of Computer Science, 2006. FOCS'06. 47th Annual IEEE Symposium on*. IEEE, 2006, pp. 667–676.

[11] T. Bu, L. Li, and R. Ramjee, "Generalized proportional fair scheduling in third generation wireless data networks," in *INFOCOM 2006. 25th IEEE International Conference on Computer Communications. Proceedings*, April 2006, pp. 1–12.

[12] Q. Ye, B. Rong, Y. Chen, M. Al-Shalash, C. Caramanis, and J. G. Andrews, "User association for load balancing in heterogeneous cellular networks," *CoRR*, vol. abs/1205.2833, 2012. [Online]. Available: http://arxiv.org/abs/1205.2833

[13] H. Kim, G. de Veciana, X. Yang, and M. Venkatachalam, "Distributed α-optimal user association and cell load balancing in wireless networks," *Networking, IEEE/ACM Transactions on*, vol. 20, no. 1, pp. 177–190, 2012.

[14] "Study on Radio Access Network (RAN) Sharing Enhancements," 3GPP TR 22.852, v12.0.0, Jun. 2013.

[15] R. Kokku, R. Mahindra, H. Zhang, and S. Rangarajan, "Nvs: A substrate for virtualizing wireless resources in cellular networks," *Networking, IEEE/ACM Transactions on*, vol. 20, no. 5, pp. 1333–1346, Oct 2012.

[16] X. Costa-Perez, J. Swetina, T. Guo, R. Mahindra, and S. Rangarajan, "Radio access network virtualization for future mobile carrier networks," *Communications Magazine, IEEE*, vol. 51, no. 7, pp. 27–35, 2013.

[17] R. Kokku, R. Mahindra, H. Zhang, and S. Rangarajan, "Cellslice: Cellular wireless resource slicing for active ran sharing," in *Communication Systems and Networks (COMSNETS), 2013 Fifth International Conference on*, 2013, pp. 1–10.

[18] C. Bernardos *et al.*, "An architecture for software defined wireless networking," *Wireless Communications, IEEE*, vol. 21, no. 3, pp. 52–61, June 2014.

[19] A. Gudipati, D. Perry, L. E. Li, and S. Katti, "Softran: Software defined radio access network," in *Proceedings of the Second ACM SIGCOMM Workshop on Hot Topics in Software Defined Networking*, ser. HotSDN '13. New York, NY, USA: ACM, 2013, pp. 25–30.

[20] R. Mahindra, M. Khojastepour, H. Zhang, and S. Rangarajan, "Radio access network sharing in cellular networks," in *Network Protocols (ICNP), 2013 21st IEEE International Conference on*, Oct 2013, pp. 1–10.

[21] A. Gudipati, L. E. Li, and S. Katti, "Radiovisor: A slicing plane for radio access networks," in *Proceedings of the Third Workshop on Hot Topics in Software Defined Networking*, ser. HotSDN '14. New York, NY, USA: ACM, 2014, pp. 237–238.

[22] A. Banchs, "User fair queuing: fair allocation of bandwidth for users," in *INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 3. IEEE, 2002, pp. 1668–1677.

[23] J. K. Lenstra, D. B. Shmoys, and E. Tardos, "Approximation algorithms for scheduling unrelated parallel machines," in *Foundations of Computer Science, 1987., 28th Annual Symposium on*. IEEE, 1987, pp. 217–224.

[24] I. Milchtaich, "Weighted congestion games with separable preferences," *Games and Economic Behavior*, vol. 67, no. 2, p. 750757, 2009.

[25] ITU-R, "Report ITU-R M.2135-1, Guidelines for evaluation of radio interface technologies for IMT-Advanced," Tech. Rep., 2009.

[26] K. Lee, S. Hong, S. J. Kim, I. Rhee, and S. Chong, "Slaw: self-similar least-action human walk," *IEEE/ACM Trans. Netw.*, vol. 20, no. 2, pp. 515–529, 2012.

APPENDIX

Corollary 1. *Let* $\sum_{p=1}^{N} w_p x_{pj}$ *be the demand at BS* j*. Then, those operators that have their users at BS with lower demand* *receive more resources from the network.*

Proof: The proof follows directly from (2). According to this equation, the throughput received by a user is inversely proportional the BS' demand. Therefore, an operator that has its users associated to BSs with lower demand will receive more throughput for its users. ■

Corollary 2. *If all the BS used by two operators have the same level of demand, resources will be distributed among them proportionally to their share.*

Proof: The proof follows directly from (2). According to this equation, for uniformly loaded base stations the throughput received by a user given by Kv_k/N_k, where K is a constant. Therefore, the total resources received by an operator that has N_k users will be Kv_k and hence proportional to v_k. ■

Corollary 3. *The amount of resources (resources blocks) that an operator receives in a base station j only depends on the fraction of total users of operator k at base station j, f_{kj}, and the share v_k.*

Proof: This can be proven directly by using theorem 2. If we let n_{kj} be the number of users from operator k in base station j, since $w_i = v_k/N_k$ and $n_{kj} = f_{kj}N_k$ the amount of resources (resources blocks) b_{kj} that an operator gets can be represented as:

$$b_{kj} = \frac{n_{kj}w_i}{n_{kj}w_i + \sum_{q\in j, q\notin k} w_q} = \frac{f_{kj}v_k}{f_{kj}v_k + \sum_{q\in j, q\notin k} w_q}$$

where it can be directly inferred that the amount of resource blocks that the operator gets only depends on the relative percent of users in the bs f_{kj} and the share v_k. ■