Joint Optimization of User Association and User Satisfaction in Heterogeneous Cellular Networks

Farah Moety, Mustapha Bouhtou, Taoufik En-Najjary, Ridha Nasri
Orange Labs Networks - Chatillon, France
Email: {farah.moety, mustapha.bouhtou, taoufik.ennajjary, ridha.nasri}@orange.com

Abstract—**In this paper, we address the joint problem of user association and resource allocation in wireless heterogeneous networks. Therefore, we formulate an optimization approach considering two objectives, namely, maximizing the number of served User Equipments (UEs) and maximizing the sum of the UE utilities. Precisely, the aim is to associate UEs with the optimal Radio Access Technology (RAT) and to allocate to these UEs the optimal Resource Units (RUs) based on their requested services and contracts. Our problem is challenging because it is mixed integer non-linear optimization. To tackle this difficulty, we provide a Mixed Integer Linear Programming (MILP) re-formulation of the problem that makes it computationally tractable. Various preferences for user association and resource allocation are conducted by tuning: on the one hand, the weights associated with different services and contracts; on the other hand, the weights associated with the considered two objectives. The optimal solution of the MILP problem is computed for a realistic network scenario and compared with legacy solution. Extensive simulation results show that the proposed optimization approach improves the overall network performance while considering the UE requested service and contract: it outperforms legacy solutions in terms of user satisfaction. Moreover, it provides an efficient distribution of UEs on the different RATs.**

I. Introduction

5G networks will consist of heterogeneous networks interoperating in a clever way offering seamless network connectivity to the user equipment (UE). In such heterogeneous network, a challenging joint problem of user association and resource allocation arises. Hence, determining which Base Station (BS) over which Radio access Technology (RAT) a given UE will be associated, and how many Resource Units (RUs) it will be allocated, all substantially affect the network performance. In the present article, we tackle the joint problem of user association and resource allocation in wireless heterogeneous networks. We thus formulate an optimization approach that jointly maximizes the number of served UEs and the sum of their utilities. This is achieved by associating UEs with the optimal BS/RAT and allocating to these UEs the optimal RUs based on their requested services and contracts. The present approach provides benefits for both UEs and network operator. Each UE is associated with the optimal RAT that guarantees its required Quality of Service (QoS) according to the contract between this UE and the network operator. The network operator in turn can better exploit its radio resources and satisfy more UEs.

In the state-of-the-art, different approaches were proposed to tackle the user association problem. First, the network-centric approach [1], where the network takes decisions transparently to UEs in a way to optimize overall network performance. Second, the user-centric approach [2], [3], where each UE selfishly strives to improve its own performance. Third, the hybrid approach [4] combines the two previous ones in such a way it does not only aim to achieve the best possible network performance, but it also considers the UE's requirement.

Authors in [1] formulated a network-centric approach for user association in a heterogeneous wireless network scenario, with various technologies and operators, as a linear optimization problem. They used a utility function with various merit parameters, reflecting the requirements of both the UEs and the network itself. In the aforementioned paper, the capacity of a BS/RAT is in terms of the number of UEs associated with this BS/RAT and not in terms of RUs. Authors in [2], [3] modeled the user association problem in wireless heterogeneous networks as a non-cooperative game, in which UEs select the BS/RATs in a distributed manner to increase their own individual throughput. Additionally, authors in [2] proposed a reinforcement learning method to find the optimal strategy of the UEs in a case where UEs have no information about other UEs. Authors in [4] formulated a hybrid approach for the user association problem. Deriving network information was formulated as a Semi-Markov Decision Process (SMDP), so as to guide UEs' decisions in a way to satisfy operator interests. Moreover, UEs combine their needs and preferences with the signaled network information, and each UE selects the BS/RAT to be associated with, in a way to maximize its own utility. However, a heavy computational load was required to find optimal solutions.

Moreover, in the literature, the problem of user association and resource allocation has not been studied jointly except in a few recent cases [5]. Authors in [5] proposed an integer linear programming model for the joint problem of user association and resource allocation with the objective of maximizing the number of associated UEs and the minimum granted utility. The authors studied the case of only one BS with multi-RATs and the UE’s utility depends only on throughput.

The key contributions of our work are as follows:

- We formulate a hybrid optimization approach that jointly maximizes the number of served UEs and the sum of their utilities. The particularity of the proposed approach that it does not only consider the network overall performance (optimal user association and optimal RUs allocation), but also, and as one of the most relevant aspect, the UEs' preferences (requested QoS and contract).
- Our problem formulation allows us to investigate various preferences for user association and resource allocation by tuning: on the one hand, the weights associated with the different service classes and contracts; on the other hand, the weights associated with the considered two objectives.

978-0-9883045-1-2/16 $31.00 © 2016 ITC

- Starting from a mixed integer non-linear formulation of the problem, we provide a Mixed Integer Linear Programming (MILP) re-formulation of the problem that makes it computationally tractable. We compute the optimal solution of the MILP problem for a realistic network scenario and compare its performance with legacy solutions.

The rest of the paper is organized as follows. In Section II, we describe the network model. In Section III we present the proposed optimization approach and the formulation of the joint problem of user association and resource allocation as a MILP problem. In Section IV we provide extensive simulation results. Conclusions and perspectives are given in Section V.

II. Network Model

We consider the downlink of a heterogeneous wireless network composed of N_{bs} BSs with N_T co-localized RATs. The indexes $i \in I=\{1,...,N_{bs}\}$, and $j \in J=\{1,...,N_T\}$, are used throughout the paper to designate a given BS and a given RAT, respectively. We term by $k \in K=\{1,...,N_u\}$, the index of a given UE where N_u is the number of UEs in the network.

A. Network Resources

In each RAT, the radio resource is divided into elementary RUs. Typically, in 4G wireless networks (*e.g.*, LTE technology), the Resource Block (RB) is the smallest RU that can be scheduled. The RB consists of 12 consecutive subcarriers for one sub-frame duration (1 ms). In 3G wireless networks (*e.g.*, HSPA technology), codes and power are treated as RUs. In this work, we consider that all codes has the same power and only codes are thus treated as RUs. Moreover, the capacity of a RAT j is limited by the number of available RUs and it is denoted by R_j.

B. Data Rate

The perceived throughput of UE k, denoted by γ_k, is the sum of the perceived throughput of this UE from the BSs/RATs of the network. Let $\varphi_{i,j,k}$ be the perceived throughput of UE k from BS i over RAT j per one RU and $\lambda_{i,j,k}$ be the number of RUs assigned to UE k associated with BS i over RAT j. When a UE is not associated with a given BS/RAT, this UE is not assigned any RU from the corresponding BS/RAT ($\lambda_{i,j,k} = 0$). Thus, γ_k is given by:

$$\gamma_k = \sum_{i \in I, j \in J} \lambda_{i,j,k}\ \varphi_{i,j,k}, \quad \forall k \in K. \tag{1}$$

Let $\nu_{i,j,k}$ be the Signal-to-Interference-plus-Noise Ratio (SINR) of UE k from BS i over RAT j, and w_j be the bandwidth per RU. Based on Shannons formula, the theoretical throughput that can be attained for UE k from BS i over RAT j per RU, is given by:

$$\varphi_{i,j,k} = w_j\ \log(1 + \nu_{i,j,k}), \quad \forall i \in I, \forall j \in J, \forall k \in K. \tag{2}$$

For instance, for LTE, w_1 is the bandwidth per one RB; for HSDPA, w_2 is the chip rate over the spreading factor.
The SINR $\nu_{i,j,k}$ is given by [6]:

$$\nu_{i,j,k} = \frac{G_t}{G_t(a+1)\mathrm{ISR}_{i,j,k} + L_{i,j,k}\frac{P_N}{P}}, \quad \forall i \in I, \forall j \in J, \forall k \in K, \tag{3}$$

where G_t is the transmit antenna gain and a is the orthogonality factor (*e.g*, $a > 0$ in 3G wireless networks, $a = 0$ in 4G networks). $L_{i,j,k}$ is the path loss detected by UE k from BS i over RAT j, P_N is the noise power and P is the power per RU, and $ISR_{i,j,k}$ is the Interference to Signal Ratio of UE k from BS i over RAT j. $ISR_{i,j,k}$ is given by [6]:

$$ISR_{i,j,k} = \sum_{i' \in I, i' \neq i} \pi_{i',j}\ \frac{L_{i,j,k}}{L_{i',j,k}}, \quad \forall i \in I, \forall j \in J, \forall k \in K, \tag{4}$$

where $\pi_{i',j}$ is the percentage of occupied resources in the interfering BS i' over RAT j.

C. Utility

We consider two types of traffic classes: non real-time and real-time classes. We use the well-known concept of utility function which maps the UE perceived throughput with the level of UE satisfaction or QoS. The index s is used throughout the paper to designate a given class of a service. Let U_k^s denote the utility function of UE k with class s service.

a) Class A traffic: includes non real-time services that are generated by traditional data applications such as file and mail download, web (*e.g.*, Twitter, Facebook), etc. These applications are assumed to be tolerant in delay variation and adapt their rate to available resource by means of a transport protocol like TCP. Thus, the elasticity of these services can be modeled by concave utility functions [7]. In this work, we assume that UEs with non real-time services have an exponential utility function given by:

$$U_k^A(\gamma_k) = 1 - e^{(-\gamma_k/\gamma^c)}, \tag{5}$$

where γ^c is the comfort throughput demand of the UE (*i.e.*, the mean throughput beyond which, UE satisfaction exceeds 63% of maximum satisfaction). The satisfaction increases slowly as the throughput exceeds the comfort throughput demand.

b) Class B traffic: includes real-time services that are generated by real-time video and voice applications. These services are partially elastic and usually characterized by a minimum, an average and a maximum data rate requirement [8]. The elasticity of these services can be modeled by a sigmoidal-like function [7]. Therefore, in this work, we consider the following sigmoidal utility function for UEs with real-time services [9]:

$$U_k^B(\gamma_k) = d_1\left(\frac{1}{1 + e^{b(\gamma^a - \gamma_k)}} - d_2\right). \tag{6}$$

where γ^a represents the average throughput demand of class B service. b is a positive constant that determines the shape of the sigmoid. $d_1 = \frac{1+e^{b\gamma^a}}{e^{b\gamma^a}}$ and $d_2 = \frac{1}{1+e^{b\gamma^a}}$.
We note that, the utility functions of the UEs with the same service class are differentiated according to the contract between these UEs and the network operator.

D. Contract

In this work, we consider that the network operator provides two differentiated types of contracts, namely, regular (R) and premium (P). They differ in their QoS, with the premium contract being the most expensive one but also the one guaranteeing higher UE satisfaction level. As a matter of fact, for the same offered average throughput, premium UEs perceive a level of satisfaction lower than that of regular UEs, as shown in Figure 1. The index t is used throughout the paper to designate a given contract.

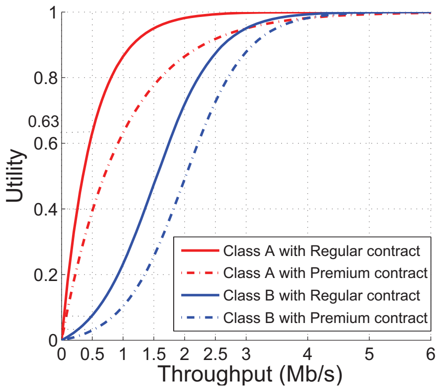


Figure 1. Utility functions for UEs with different service class and contract.

III. Optimization Problem

A. Problem Formulation

The proposed optimization approach consists of finding an optimal user association and an optimal RUs allocation that jointly maximize the number of served UEs and the sum of their utilities. It takes into account the required QoS of the UE's requested service and the contract between this UE and the network operator. Let $K^{s,t}$ denote the set of UEs with class s service and contract t.

The design variables in our maximization problem are as follows:

- The user association with the network BSs over a given RAT.
- The number of RUs assigned to a given UE associated with a given BS/RAT.

Let Θ be the matrix, with elements $\theta_{i,j,k}$, defining the user association with the network BSs over a given RAT; and $\theta_{i,j,k}$ be a binary variable that indicates whether or not UE k is associated with BS i over RAT j.

$$\theta_{i,j,k} = \begin{cases} 1 & \text{if UE } k \text{ is associated with BS } i \text{ over RAT } j, \\ 0 & \text{otherwise.} \end{cases}$$

Let Λ be the matrix, with elements $\lambda_{i,j,k}$, defining the amount of RUs allocated to a given UE from a given BS/RAT. $\lambda_{i,j,k}$ is an integer variable that indicates the number or RUs assigned to UE k associated with BS i over RAT j.

In the following, we define the constraints on the decision variables and the utility functions. We start by defining the constraints on the user association and the RUs as follows:

$$\sum_{i \in I, j \in J} \theta_{i,j,k} \leq 1, \quad \forall k \in K. \tag{7}$$

$$\sum_{k \in K} \lambda_{i,j,k} \leq R_j, \quad \forall i \in I, \forall j \in J. \tag{8}$$

$$\lambda_{i,j,k} \geq \theta_{i,j,k}, \quad \forall i \in I, \forall j \in J, \forall k \in K. \tag{9}$$

$$\lambda_{i,j,k} \leq R_j \ \theta_{i,j,k}, \quad \forall i \in I, \forall j \in J, \forall k \in K. \tag{10}$$

Constraints (7) state that a given UE can be associated with only one BS over one RAT. Constraints (8) ensure that the limit on the number of RUs for each RAT, R_j, is not exceeded. Constraints (9) ensure that $\lambda_{i,j,k}$ is not equal to zero when $\theta_{i,j,k}$ is equal to one. When a UE is associated with a given BS/RAT, these constraints ensure that this UE is assigned a number of RUs from the corresponding BS/RAT. Constraints (10) force $\lambda_{i,j,k}$ to be equal to zero when $\theta_{i,j,k}$ is equal to zero. When a UE is not associated with a given BS/RAT, these constraints prevent this UE from being assigned a number of RUs from the corresponding BS/RAT.

Let us introduce some notations to define the constraints on the utility functions. Let $u^{min,s,t}$ be the minimum required utility for class s service with contract t, and $u^{max,s,t}$ be the maximum utility that the operator is willing to offer for service s with contract t. The constraints on the utility functions are as follows:

$$U_k^{s,t} \geq u^{min,s,t} \ \theta_{i,j,k}, \quad \forall i \in I, \forall j \in J, \forall k \in K^{s,t}, \ \forall s \in \{A, B\}, \forall t \in \{P, R\}, \tag{11}$$

$$U_k^{s,t} \leq u^{max,s,t} \ \theta_{i,j,k}, \quad \forall i \in I, \forall j \in J, \forall k \in K^{s,t}, \ \forall s \in \{A, B\}, \forall t \in \{P, R\}, \tag{12}$$

In case a given UE is associated with a given BS/RAT, constraints (11) ensure that the minimum required level of satisfaction for this UE according to its service and contract, is guaranteed. In other words, if the minimum required level of satisfaction for a given UE cannot be guaranteed by a given BS/RAT, this UE will not be associated with this BS/RAT to preserve the overall network performance.

In case a given UE is associated with a given BS/RAT, constraints (12) ensure that the maximum level of satisfaction for this UE according to its service and contract, is not exceeded. In fact, we add these constraints to ensure an effective use of the RUs available in the network. For instance, for class A service, as the utility function is concave, the satisfaction increases slowly as the throughput exceeds the comfort throughput demand (as shown in figure 1). Thus, limiting the satisfaction to a maximum level, prevents, some UEs from being assigned relatively a high number of RUs without increasing considerably their satisfaction. Similarly, for class B service, the utility function is a sigmoidal-like function, which is concave for a throughput higher than the average throughput demand.

Let $\gamma^{c,A,t}$ denote the comfort throughput demand for class A service and contract t, and $\gamma^{a,B,t}$ denote the average throughput demand for class B service and contract t. The expressions of the utility functions of UEs with class A and B services, and contract t, are respectively given by:

$$U_k^{s,t}(\gamma_k) = 1 - e^{(-\frac{\gamma_k}{\gamma^{c,s,t}})}, \quad \forall k \in K^{A,t}, \forall s \in \{A\}, \ \forall t \in \{P, R\}, \tag{13}$$

$$U_k^{s,t}(\gamma_k) = d_1^{s,t}\left(\frac{1}{1 + e^{b(\gamma^{a,s,t} - \gamma_k)}} - d_2^{s,t}\right), \quad \forall k \in K^{B,t}, \ \forall s \in \{B\}, \forall t \in \{P, R\}, \tag{14}$$

where $d_1^{s,t} = \frac{1 + e^{b\gamma^{a,s,t}}}{e^{b\gamma^{a,s,t}}}$ and $d_2^{s,t} = \frac{1}{1 + e^{b\gamma^{a,s,t}}}, \forall s \in \{B\}, \forall t \in \{P, R\}$.

The integrality constraints for the decision variables $\theta_{i,j,k}$ and $\lambda_{i,j,k}$ are respectively given by:

$$\theta_{i,j,k} \in \{0, 1\}, \quad \forall i \in I, \forall j \in J, \forall k \in K, \tag{15}$$

$$\lambda_{i,j,k} \in \mathbb{N}, \quad \forall i \in I, \forall j \in J, \forall k \in K. \tag{16}$$

To eliminate some trivial cases that must not be included in the solution, we add the following constraints:
If UE k is not covered by BS i over RAT j, then

$$\theta_{i,j,k} = 0. \tag{17}$$

The equation (17) prevents a given UE from being associated with a BS/RAT if this UE is not in its coverage area. We note that the coverage area of a given BS/RAT is defined as the geographical area where the received SINR of each UE is above a given minimum threshold.
Therefore, our approach can be formulated as an optimization problem $(\mathcal{P})$ that consists of jointly maximizing the weighted sum of the number of served UEs in the network and their utilities subject to the aforementioned constraints. Consequently, problem $(\mathcal{P})$ is given by:

$$\text{Maximize} \quad \sum_{\substack{k\in K^{s,t},\\ s\in\{A,B\},\\ t\in\{P,R\}}} \alpha^{s,t}(\beta_1 \sum_{\substack{i\in I,\\ j\in J}} \frac{\theta_{i,j,k}}{|K^{s,t}|} + \beta_2 U_k^{s,t}), \tag{18}$$
$$\text{subject to:} \quad (1)\text{and}(7)\text{to}(17),$$

where $\alpha^{s,t}$ are the weighting factors corresponding to the UEs with different services and contracts. Tuning these factors, allows to privilege some UEs based on their service classes and contracts. This in turns allows to consider various preferences for user association and RUs allocation. It is usually assumed that $\sum_{s\in\{A,B\},t\in\{P,R\}} \alpha^{s,t}$=1, and that $\alpha^{s,t} \in$ [0,1]. In particular, when $\alpha^{B,P} > \alpha^{B,R} > \alpha^{A,P} > \alpha^{A,R}$, UEs with class B service and premium contract are the most privileged UEs, followed by UEs with class B service and regular contract, then UEs with class A service and premium contract, and finally UEs with class A service and regular contract have the lowest privilege.
Similarly, β_1 and β_2 are the weighting factors representing the relative importance of the two objectives (namely, the number of served UEs, and the sum of UE utilities). Moreover, $\beta_1 + \beta_2 = 1$, β_1 and $\beta_2 \in$ [0,1]. In particular, when β_1 equals 1 and β_2 equals 0, we only focus on the maximizing the number of served UEs, and as β_1 decreases and β_2 increases more importance is given on the maximization of the sum of UE utilities.

B. From non-linear to linear optimization problem

Problem $(\mathcal{P})$ is a non-linear mixed integer optimization problem. The non-linearity comes from the expression of the UE utility functions (exponential and sigmoidal functions). Solving such problem is a very challenging task. In this section, we explain how to transform problem $(\mathcal{P})$ into a MILP problem. A MILP problem consists of a linear objective function, a set of linear equality and inequality constraints and a set of variables with integer restrictions. Generally, MILP problems are solved using a linear-programming based branch-and-bound approach [10]. The idea of this approach is to solve Linear Program (LP) relaxations of the MILP and to look for an integer solution by branching and bounding on the decision variables provided by the LP relaxations.

1) Methodology: Equations (1) show that the possible values of UE's throughput are discrete and not continuous. This implies that the possible values of UE's utility are also discrete, which allows us to transform problem $(\mathcal{P})$ into a MILP problem $(\mathcal{P}_1)$.

Let us introduce some notation to describe the MILP reformulation. Let $n \in \mathcal{R}^j$={1, ..., R_j}, be the possible number of RUs that a UE can obtain when associated with a given BS over RAT j. Let $\eta^j_{i,k,n}$, be the throughput of UE k, associated with BS i over RAT j and assigned n RUs. Let $\mu^{j,s,t}_{i,k,n}$, be the utility of UE k, with class s service and contract t, associated with BS i over RAT j and assigned n RUs. First, we compute all the possible values of the throughput a UE can obtain from the different BS/RATs (*i.e*, the values of $\eta^j_{i,k,n}$). Second, similarly, we compute all the possible values of the UE's utility from the different BS/RATs (*i.e*, the values of $\mu^{j,s,t}_{i,k,n}$).
Once we know all the possible values of the UE's utility from the different BS/RATs, we reformulate the considered problem of user association and RUs allocation as a linear knapsack problem, where RUs of a given UE are the objects to be chosen and the BS/RATs are the knapsacks in such a way to: (i) jointly maximize the number of served UEs and the sum of their utilities; (ii) take into account the required QoS of the UE's requested service and the contract between this UE and the network operator.
With the linear Knapsack reformulation, we thus introduce a new binary variable $x^j_{i,k,n}$, equaling to one if n RUs are assigned to UE k associated with BS i over RAT j, and zero otherwise. Therefore, $\lambda_{i,j,k}$ is given by:

$$\lambda_{i,j,k} = \sum_{n\in\mathcal{R}^j} n\, x^j_{i,k,n}, \quad \forall i \in I, \forall j \in J, \forall k \in K. \tag{19}$$

Consequently, our optimization problem is formulated as a MILP problem $(\mathcal{P}_1)$ and it is given by:

$$\text{Maximize} \quad \sum_{\substack{k\in K^{s,t},\\ s\in\{A,B\},t\in\{P,R\}}} \alpha^{s,t}(\beta_1 \sum_{\substack{i\in I,\\ j\in J}} \frac{\theta_{i,j,k}}{|K^{s,t}|} +$$
$$\beta_2 \sum_{i\in I,j\in J,n\in\mathcal{R}^j} x^j_{i,k,n}\, \mu^{j,s,t}_{i,k,n}), \tag{20}$$
$$\text{subject to:} \quad (7)-(10),(15)-(17)\text{and}(19),$$

$$\sum_{i\in I,j\in J,n\in\mathcal{R}^j} x^j_{i,k,n} \le 1, \quad \forall k \in K, \tag{21}$$

$$x^j_{i,k,n} \le \theta_{i,j,k}, \quad \forall i \in I, \forall j \in J, \forall k \in K, \forall n \in \mathcal{R}^j, \tag{22}$$

$$\sum_{n\in\mathcal{R}^j} x^j_{i,k,n}\, \mu^{j,s,t}_{i,k,n} \ge u^{min,s,t}\, \theta_{i,j,k}, \quad \forall i \in I, \forall j \in J, \forall k \in K^{s,t}, \forall s \in \{A,B\}, \forall t \in \{P,R\}, \tag{23}$$

$$\sum_{n\in\mathcal{R}^j} x^j_{i,k,n}\, \mu^{j,s,t}_{i,k,n} \le u^{max,s,t}\, \theta_{i,j,k}, \quad \forall i \in I, \forall j \in J, \forall k \in K^{s,t}, \forall s \in \{A,B\}, \forall t \in \{P,R\}, \tag{24}$$

$$x^j_{i,k,n} \in \{0,1\}, \quad \forall i \in I, \forall j \in J, \forall k \in K, \forall n \in \mathcal{R}^j. \tag{25}$$

Problem $(\mathcal{P}_1)$ is equivalent to problem $(\mathcal{P})$. Precisely, constraints (21) state that a given UE can be assigned a given number of RUs from only one BS over one RAT. Constraints

(22) force $x^j_{i,k,n}$ to be equal to zero when $\theta_{i,j,k}$ is equal to zero. These constraints prevent a UE from being assigned a number of RUs if he is not associated with a given BS/RAT. In fact, constraints (22) are called valid inequalities as they help us to find the optimal solution. Constraints (23) and constraints (24) replace constraints (11) and (12), respectively. Finally, constraints (25) are the integrality constraints for the variable $x^j_{i,k,n}$.

IV. Performance Evaluation

A. Evaluation Methodology

We consider the realistic positioning of the 4G and 3G network BSs for the district 14 of Paris-France [11]. The network topology is composed of 18 cells (N_{bs}=18) with two co-localized RATs, namely LTE and HSDPA[1]. The positioning of UEs follows a random uniform distribution as shown in Fig. 2. For simplicity, we assume that each considered cell has

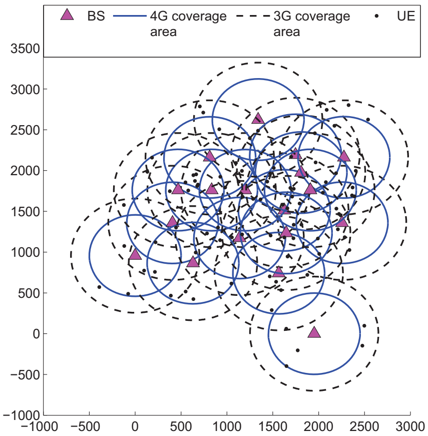


Figure 2. Network topology of the 14th district of Paris with 90 UEs in the network.

an Omni-directional radiation pattern. We also assume that all interfering BSs has the same percentage of occupied resources and thus $\pi_{i',j} = \pi_j$ (*cf.* Eq. (4)). The simulated LTE system bandwidth is 10 MHz, therefore we have 50 RBs available in each cell. We assume that two RBs are used for special information (*e.g*, signaling, etc.), and we thus have 48 RBs available in each cell. For HSDPA, the system bandwidth is 5 MHz, therefore we have 16 codes available in each cell. We assume that two codes are reserved for special information (*e.g*, signaling, etc.), and we thus have 14 codes available in each cell. The simulation parameters and the pathloss model follow that in [12], which are summarized in Tab. I.

The path loss between the BS/RAT and the UE is computed according to the Cost 231 extended Hata model considering a urban environment [12], with a carrier frequency f_1 of 2600 MHz for 4G and a carrier frequency f_2 of 2100 MHz for 3G. The shadowing [dB scale] is represented by a random variable following normal distribution with a mean of 0 dB and a standard deviation of 10 dB. The coverage radius is 500 m for LTE and 700 m for HSDPA. Table II shows how much UEs are covered by the considered BS/RATs for the

[1]We will use the terms RAT 1, 4G or LTE interchangeably throughout the paper. Similarly, for the terms RAT 2, 3G or HSDPA.

Table I
Simulation parameters for LTE and HSDPA

Parameter	**Value**	
	RAT 1 (4G)	RAT 2 (3G)
Coverage radius	500 m	700 m
Carrier frequency	2600 MHz	2100 MHz
Bandwidth	10 MHz	5 MHz
Number of RUs per cell	R_1 = 48	R_2 = 14
Orthogonality factor (a) (*cf.* Eq. (3))	0	0.5
Percentage of occupied resources (π_j)	80%	90%
Transmit power	10 W	
Antenna	Omni-directional	
Transmit antenna gain (G_t) (*cf.* Eq. (3))	15 dBi	
Noise Figure	9 dB	
Environment	Urban	
Pathloss model	Cost 231 extended Hata model	
Shadowing standard deviation	10 dB	

considered number of UEs in the network in these simulations. For instance, for 90 UEs in the network, 21 UEs are covered by RAT 2, where only 11 of them are also covered by RAT 1. These 11 UEs have the choice to be associated with one of the two RATs or blocked, while the remaining 10 UEs can only be associated with RAT 2 or blocked.

Table III shows the simulation parameters of UEs with differ-

Table II
Average Number of UEs in the coverage area of the BS/RATs.

Number of UEs in the network	90	180	270	360	540	720
RAT 1 = 4G	11	22	33	45	66	88
RAT 2 = 3G	21	42	64	86	128	171

ent service classes and contracts. First, it shows the percentage of these UEs in the network. Second, it depicts the QoS demands of UEs regarding their service classes and contracts. QoS demands are expressed in terms of: i) comfort throughput $\gamma^{c,A,t}$ (cf. Eq(13)) for UEs with class A service and contract t, and ii) average throughput $\gamma^{a,B,t}$ (cf. Eq(14)) for UEs with class B service and contract t.

Table IV shows the values of the minimum required utility

Table III
Simulation parameters for UEs with different service classes and contracts.

	Class A	Class B
	Percentage of UEs	
Regular	40%	10%
Premium	20%	30%
	Throughput demand of UEs	
Regular	$\gamma^{c,A,R}$=1 Mb/s	$\gamma^{a,B,R}$=3 Mb/s
Premium	$\gamma^{c,A,P}$=2 Mb/s	$\gamma^{a,B,P}$=4 Mb/s

($u^{min,s,t}$, *cf.* constraints (23)) and the values of the maximum utility ($u^{max,s,t}$, *cf.* constraints (24)) for each service class and contract used in our simulations. Indeed, since UEs with class A service adapt to resource availability, and require no QoS guarantees, their minimum utility requirement is zero ($u^{min,A,R/P} = 0$). However, UEs with class B service are characterized by a minimum data rate requirement; decreasing below a certain threshold will result in a drop in the QoS. Therefore, we consider that the minimum required level of utility to be guaranteed, is equal to 10% ($u^{min,B,R/P} = 10\%$). As mentioned in Section III-A, limiting the satisfaction to a maximum level, prevents some UEs from being assigned a relatively high number of RUs without increasing considerably their satisfaction. For UEs with class A service, the maximum utility is equal to $u^{max,A,R/P} = 77.70\%$, which corresponds to an average throughput equal to 1.5 times the comfort throughput demand. We consider that UEs with class B service are more privileged than UEs with class A service, thus their maximum utility is higher and it is equal to $u^{max,B,R/P} = 95\%$.

Table IV
Minimum and Maximum utility for each service class and contract.

	$u^{min,s,t}$	$u^{max,s,t}$
Class A Regular/Premium	0%	77.70%
Class B Regular/Premium	10%	95%

Toward studying the performance of the proposed approach, we investigate the optimal solutions obtained by tuning: on the one hand, the weights associated with different service classes and contracts; on the other hand, the weights associated with the considered two objectives (maximizing the number of served UEs and maximizing the sum of the UE utilities). We thus consider three settings illustrated in Table V. Setting S1 matches the case where all UEs are equally important in the optimization, and both objectives are equally important. Setting S2 matches the case where UEs with class B service and premium contract are the most privileged UEs, followed by UEs with class B service and regular contract, then UEs with class A service and premium contract, and finally UEs with class A service and regular contract have the lowest privilege. Moreover, both objectives are equally important in this setting. Finally, setting S3 is similar to S2 in the privilege of the different service classes and contracts. However, it is different from other settings in the relative importance of the two objectives, where more importance is given to maximizing the number of served UEs.

We compute the optimal solution of the MILP problem $(\mathcal{P}1)$ using the CPLEX V12.6.0.0 solver running on a computer equipped with an Intel(R) Xeon(R) CPU L5630, 4 cores, and a clock rate of 2.13 GHz. This tool provides the optimal solution using the branch and cut approach [13] (which consists of a combination of a cutting plane method with a branch-and-bound algorithm). The solver configuration used is the default setting. All the solutions are thus provided at the optimum. The input data for the CPLEX solver are generated using MATLAB. Thus, in MATLAB, we implement the considered heterogeneous network topology of the 14th district of Paris. We adopt the Monte Carlo method by generating 10 snapshots with different random uniform UEs distribution. After doing the calculations for all the snapshots, we provide the 95% Confidence Interval (CI) for each simulation result. We compare the performance of the MILP solutions for the considered settings with an existing approach for user association and resource allocation presented in the sequel.

B. Existing Approach

In legacy cellular networks, UEs are associated with the BS/RAT delivering the Highest Received Power (HRP) of pilot signals [14] with a prioritization for the last network generation. Moreover, RUs are shared equitably between UEs meaning that all UEs get a similar number of RUs.

In the present paper, we devise a reference model denoted by HRP and based on legacy networks considering a higher priority association for RAT 1 (4G). The reference model HRP works as follows:

- The UE measures the received power of pilot signals for the covering BSs/RATs, and keeps them in a descendent order queue for each RAT.
- Starting with RAT 1 queue, the UE is associated with the first BS/RAT 1 that positively confirmed its request among the list in this queue. A UE request may be blocked for different reasons like limitations on the capacity at the BS/RAT 1 and the minimum required received power (-85dbm [15]).
- In case of a rejection from all BSs in the RAT 1 queue, the UE repeats the same procedure for RAT 2 queue. Similarly, a UE request may be rejected for limitations on the resource capacity.

Once the association is done with a given BS/RAT, the UE shares equitably the RUs with other UEs associated with the same BS/RAT.

C. Simulation Results

We now investigate the optimal solutions obtained by the considered settings, and the solution of HRP. First, we start by evaluating the percentage of served UEs. Second, we present the user satisfaction, which is expressed by the average utility per UE per service class per contract. Finally, we show how the considered optimization approach load balances the UEs on the different RATs.

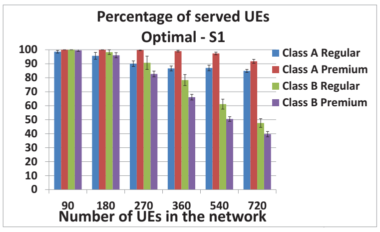
1) Percentage of served UEs: Figure 3 shows the percentage of served UEs with different service classes and contracts as a function of the number of UEs in the network, for the solutions of the considered settings. Intuitively, as the number of UEs in the network increases, the percentage of served UEs decreases because of the limitations on the available RUs at the BSs/RATs.

For the case where all UEs and both objectives are equally important in the optimization, Figure 3(a) shows that the percentage of served UEs with class A service (both regular and premium contracts) is higher than that of UEs with class B service (both regular and premium contracts). In fact, the objective function of problem $(\mathcal{P}_1)$ consists of jointly maximizing the number of served UEs and the sum of their utilities. Moreover, for the same offered throughput, UEs with class A service have a higher utility compared with UEs with class B service (as shown in Figure 1). Therefore, UEs with class A service are much more accepted than others in this setting.

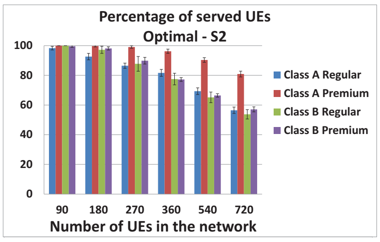
Setting S2 overcomes this issue by privileging UEs with class B service. Precisely, Figure 3(b) shows that the percentage of served UEs with class B service (both premium and regular contracts) has increased, while the percentage of served UEs with class A service (both premium and regular contracts) has decreased. We note that, although the percentage of served UEs with class B service (both premium and regular) has increased in setting S2, this percentage is still relatively low for relatively high number of UEs in the network. This is because, for the same level of satisfaction, UEs with class B service are more exigent in throughput than UEs with class A service (as shown in Figure 1). Moreover, a UE with class B service is served only if it will be allocated a given number of RUs

Table V
Three studied settings.

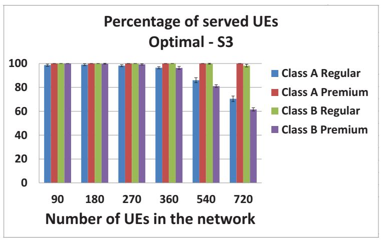
Settings	Weighting coefficients value associated with different service classes and contracts	objectives (number of served UEs and sum of UEs utilities)
S1	$\alpha^{B,P} = \alpha^{B,R} = \alpha^{A,P} = \alpha^{A,R}$=0.25	$\beta_1 = \beta_2$=0.5
S2	$\alpha^{B,P}$=0.4, $\alpha^{B,R}$=0.3, $\alpha^{A,P}$=0.2, $\alpha^{A,R}$=0.1	$\beta_1 = \beta_2$=0.5
S3	$\alpha^{B,P}$=0.4, $\alpha^{B,R}$=0.3, $\alpha^{A,P}$=0.2, $\alpha^{A,R}$=0.1	β_1=0.99, β_2=0.01



(a) Optimal-S1.



(b) Optimal-S2.



(c) Optimal-S3.

Figure 3. Percentage of served UEs for the solutions of the considered settings. (a) (b) (c).

that satisfies its minimum required utility (*cf.* constraint 23), otherwise this UE will be blocked.

Since more importance in setting S3 is given to maximizing the number of served UEs, Figure 3(c) shows that in this setting the percentage of served UEs with different service classes and contracts has increased compared with the two previous settings. Yet, for a relatively high number of UEs in the network (*e.g.*, $\geq$ 540), the network operator can deploy small cells or WiFi access points to accommodate the remaining UEs that could not be served in the macro cells.

We note that, in HRP model, all UEs with different service classes and contracts are served, except for a very low percentage of UEs ($\leq$ 1%) that are blocked for a high number of UEs in the network. This is because, in HRP model, the UE' s demands are not taken into account. Thus, the HRP model accepts UEs and shares equitably the RUs between them as long as the limit on the capacity (in terms of RUS) at the BS/RAT is not exceeded, and the minimum required received power for association is verified. However, in the next results, we show how the solutions of the proposed optimization approach leverage the tradeoff between enhancing the network overall performance while satisfying the UE' s demands and blocking some UEs.

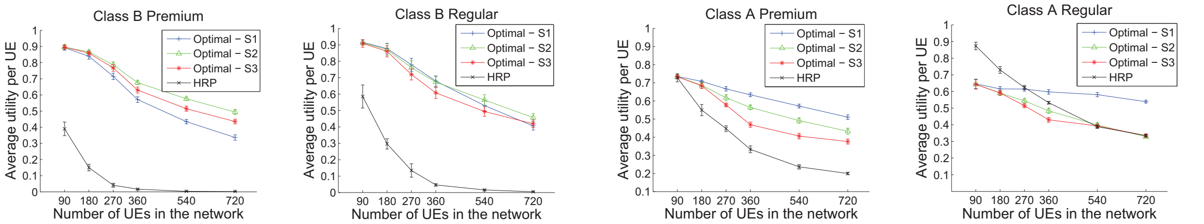
2) User satisfaction: Figure 4 shows the average utility per UE for each service class and contract, for the solutions of the considered settings and for the reference model HRP. For all the solutions, the average utility per UE per service per contract has a decreasing function. This is because, as the number of UEs in the network increases, the UEs will be allocated a lower number of RUs, which lower their achieved throughput and thereby lower their average utility. Moreover, for the solutions of the considered settings, this is also because of the percentage of served UEs, which is a decreasing function of the number of UEs in the network (as shown in Figure 3). The blocked UEs, having a utility equals zero, contribute in reducing on average the utility per UE in the network.

For UEs with class B service and premium contract, Figure 4(a) shows that the optimal solutions of the considered settings S1, S2 and S3 outperform HRP. For a low number of UEs in the network, S1, S2 and S3 solutions have relatively same performance. As the number of UEs in the network increases, the gap between the these three solutions increases, with S2 having the highest average utility per UE, followed by S3 then S1. This is because, UEs with class B service and premium contract are the most more privileged in setting S2 and S3. This increases the number of served UEs with class B service and premium contract for solutions S2 and S3 (as shown in Figure 3), and thereby increases the average utility per UE. Yet, in setting S3, more importance is given to serve UEs than to maximize their utilities. This causes the average utility per UE for solution S3 to be lower than that of solution S2.

Moreover, we noticed that the performance of HRP solution degrades drastically with the increase of the number of UEs in the network. In particular, the average utility per UE with class B service and premium contract degrades from 39% to 0.20%, when the number of UEs in the network increases from 90 to 720. In fact, HRP model treats equally all the UEs in the network. With a high number of UEs in the network, each UE will thus be allocated a relatively low number of RUs. This decreases the achievable throughput for all UEs; decreasing below a certain threshold, the satisfaction of UEs with class B service will drop drastically (as shown in Figure 1). Compared to all other type of UEs, HRP solution gives the lowest average utility per UE with class B service and premium contract.

For UEs with class B service and regular contract, Figure 4(b) shows that the optimal solutions of the considered settings S1, S2 and S3 outperform HRP. Precisely, S1 and S2 solutions have relatively same performance and an average utility per UE greater than S3. This is explained by the fact that more importance in S3 is given to serve UEs than to maximize their utilities.

Figure 4(c) shows the average utility per UE with class A service and premium contract. For relatively low number of UEs in the network, all solutions have the same performance. As the number of UEs in the network increases, S1 shows the highest average utility per UE followed by S2, then S3, and finally HRP. As UEs with class A service and premium contract have low privilege in S2 and S3, their average utilities



(a) Class B service and premium contract (b) Class B service and regular contract (c) Class A service and premium contract (d) Class A service and regular contract

Figure 4. Average utility per UE per service per contract (a) (b) (c) (d).

are lower than that of S1.

Figure 4(d) shows the average utility per UE with class A service and regular contract. For a number of UEs in the network less then 270, HRP solution has the highest average utility per UE followed by S1, S2, then S3. Between 270 and 540 UEs in the network, S1 has the highest average utility per UE followed by HRP, then S2 and finally S3. For a number of UEs in the network greater then 540, S1 has the highest average utility per UE and S2, S3 and HRP have same performance. In fact, in HRP model, RUS are equitably shared between UEs, thus the lower the number of UEs in the network, the higher the number of RUs allocated to each UE, thus the higher the average utility per UE. Moreover, for the same offered throughput, UEs with class A service and regular contract has the highest utility compared with other UEs (*cf.* Figure 1). We note that, S2 and S3 have relatively the lowest average utility per UE with class A service and regular contract, because these UEs have the lowest privilege in these settings. Moreover, the gap between these two solutions and S1 solution increases with the increase of UEs in the network. This is due to the decrease in the number of served UEs with class A service and regular contract for solutions S2 and S3 (as shown in Figure 3).

We note that, in our studied simulations, we limited the satisfaction of UEs to 77.70% for class A service and to 95% for class B service. The effect of these limitations (cf. constraints (24)) on the maximum utility can be clearly seen. For instance, the maximum achievable utility per UE with class A service and regular contract is equal to 64% for the three settings (as shown in Figure 4(d)). Moreover, the present results show that lowering the maximum satisfaction for UEs with class A service, plays a role in enhancing the satisfaction of UEs with class B service. Particularly, for 90 UEs in the network, for S3, the average utility per UE equals 64% for UEs with class A service and regular contract (see Figure 4(d)) and it equals 90% for UEs with class B service and premium contract (see Figure 4(a)). However, for HRP solution, the former is 87.34% (see Figure 4(d)) and the latter is 39% (see Figure 4(a)). This reveals the importance of our optimization approach that takes into consideration the UE's requested service and the contract between this UE and the network operator.

In conclusion, by tuning the weighting coefficients, we obtain different points located on the Pareto frontier presenting all the compromises between the satisfaction of UEs with different services and contracts on the one hand, and the two objectives on the other hand. The network operator can thus fine-tune the model to reflect its own decision preferences. For instance, the operator has the choice to privilege the satisfaction of some UEs based on their services and contracts, to give more importance for maximizing the number of served UEs, or to balance the tradeoff between maximizing the number of served UEs and maximizing their utilities.

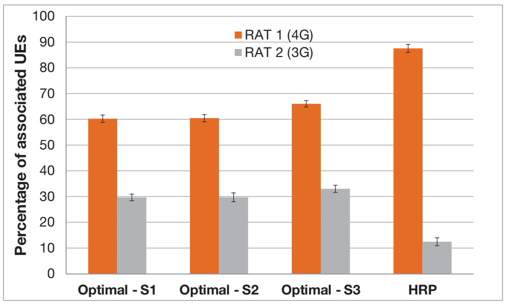


Figure 5. Percentage of UEs associated with 4G and 3G for 270 UEs in the network.

3) UEs distribution on different RATs: Figure 5 shows the percentage of UEs associated per BS over both RATs, for the solutions of the considered settings and for HRP solution, for the case of 270 UEs in the network. For the solution of the considered settings, UEs are efficiently distributed between the two available RATs. Whereas, for HRP solution, the majority of UEs are associated with RAT 1. For instance, in S3, 66% of UEs are associated with RAT 1 and 33% of UEs are associated with RAT 2. However, in HRP solution, 88% UEs associated are with RAT 1 and 22% are associated with RAT 2. In fact, in HRP solution, UEs are associated with the BS/RAT delivering the highest received power of pilot signals with a prioritization for RAT 1. Thus, covered by both RATs, UEs are more likely associated with the prioritized RAT (which is RAT 1) as long as the minimum required level of received power is verified. This causes the rush on RAT 1, and thereby reduces the average utility per UE, specially for a relatively high number of UEs int he network.

D. Discussion

In this study, we assume the existence of a central entity (CE) that has a complete control of the network state and elements (such as UEs and BSs/RATs). This entity can be easily introduced to the current wireless access networks [16]. The CE senses the network state information such as radio channel conditions, QoS demand of UEs, etc. After collecting the necessary information (*e.g.*, using the IEEE 1900.4 standard [17]), the CE intervenes at regular time intervals (periodic intervention) and provides the optimal solution of the proposed optimization problem. Precisely, the CE guides the UEs to be associated with the appropriate BS/RAT. Moreover, it diffuses to the network BSs/RATs their optimal RUs allocation.

In fact, the computation time of the optimal solution has a great impact on the periodic intervention of the CE. In the present simulations, the optimal solutions of the different settings are obtained using CPLEX solver. All the solutions were provided for a gap-to-optimality equals zero. The gap-to-optimality metric expresses the gap between the obtained integer solution and the optimal solution estimated by the solver. We noticed that the proposed optimization approach has a reasonable computation time. Particularly, the computation time of the optimal solutions varies on average between 1s and 26s depending on the number of UEs in the network and the considered setting. Therefore, the present study brings a value in providing an optimal solution of the challenging joint problem of user association and resource allocation in heterogeneous networks that has low computational complexity for a realistic network scenario.

We now investigate the impact of varying the gap-to-optimality metric on the computation time of the obtained solution. We consider the solution of setting S3 where we have 180 UEs in the network. Table VI shows that the mean computation time of setting S3 solution decreases as the gap-to-optionality increases. In particular, for a mean gap to optimality equals 0%, the mean computation time equals 17.86 s, and with an increase of the gap to optimality to 4.13%, the mean computation time decreases to 1.68 s. Therefore, a network operator has the option to choose the operation point of the network. For instance, the operator can choose the gap-to-optimality, that provides a near-optimal solution within a very short time. This also helps in reducing the interval time of the periodic intervention of the CE. In future work, more advanced techniques in the optimization process will be applied to further reduce the computation time of the optimal solution.

Table VI
Gap-to-optimality Vs computation time for 180 UEs in the network for S3.

Mean gap-to-optimality [%]	0	0.92	1.62	4.13
Mean computation time [s]	17.86	2.06	1.95	1.68

V. Conclusion

In this paper, we proposed a hybrid approach with the aim of optimizing not only the overall network performance and but also the UE satisfaction regarding its QoS demand and contract. We thus formulate an optimization problem considering double objectives, namely, maximizing the number of served UEs and maximizing the sum of the UE utilities. Starting from a mixed integer non-linear formulation of the problem, we provide a MILP reformulation of the problem that makes it computationally tractable. Different settings reflecting various preferences were carried out by tuning on the one hand, the weights associated with different services and contracts; on the other hand, the weights associated with the considered two objectives. The optimal solution of the MILP problem is computed for a realistic network scenario and compared with legacy solution. The extensive simulation results show that the proposed approach outperforms legacy solution in terms of user satisfaction. For instance, when adequately tuned, for a relatively high number of UEs in the network, the proposed approach offers for UEs with real-time service (class B) and premium contract an average utility of 50%, whereas the existing solution can not offer them more than 0.20% of satisfaction. Moreover, it leverages the tradeoff between distributing UEs on the different RATs and blocking some others, in a way that enhances the overall performance of the network. Furthermore, the present study brings a value in providing an optimal solution of this challenging problem that has a low computational complexity for realistic network scenarios. For future work, we plan to study the dynamics of the network. In particular, we need to take into consideration the mobility of UEs, the arrival and departure of the UEs in the network. Moreover, we plan to examine heuristic algorithms to solve this challenging problem.

References

[1] Johnny Choque, Ramón Agüero, and Luis Muñoz. Optimum selection of access networks within heterogeneous wireless environments based on linear programming techniques. *Mobile Networks and Applications*, 16(4):412–423, 2011.

[2] P. Naghavi, H. Rastegar, V. Shah-Mansouri, and H. Kebriaei. Learning rat selection game in 5g heterogeneous networks. *IEEE Wireless Communications Letters*, PP(99):1–1, 2015.

[3] Ehsan Aryafar, Alireza Keshavarz-Haddad, Michael Wang, and Mung Chiang. Rat selection games in hetnets. In *Proc. IEE INFOCOM*, pages 998–1006, 2013.

[4] M. El Helou, M. Ibrahim, S. Lahoud, K. Khawam, D. Mezher, and B. Cousin. A network-assisted approach for rat selection in heterogeneous cellular networks. *IEEE Journal on Selected Areas in Communications*, 33(6):1055–1067, 2015.

[5] Anna Zakrzewska, Fabio D'Andreagiovanni, Sarah Ruepp, and Michael S Berger. Biobjective optimization of radio access technology selection and resource allocation in heterogeneous wireless networks. In *Proc. 11th International Symposium on Modeling & Optimization in Mobile, Ad Hoc & Wireless Networks (WiOpt)*, pages 652–658, 2013.

[6] R. Nasri and A. Jaziri. On the analytical tractability of hexagonal network model with random user location. *IEEE Transactions on Wireless Communications*, 15(5):3768–3780, 2016.

[7] Jang won Lee, Ravi R. Mazumdar, and Ness B. Shroff. Non-convex optimization and rate control for multi-class services in the internet. *IEEE/ACM Transactions on Networking*, 13:827–840, 2005.

[8] 3GPP. Quality of service (qos) concept and architecture (release 6). TS 23.107, 3rd Generation Partnership Project (3GPP), 2005-06.

[9] Q. T. Nguyen-Vuong, Y. Ghamri-Doudane, and N. Agoulmine. On utility models for access network selection in wireless heterogeneous networks. In *Proc. of IEEE Network Operations and Management Symposium (NOMS)*, pages 144–151, 2008.

[10] J. Clausen. Branch and bound algorithmsprinciples and examples, 1999. Department of computer science, University of Copenhagen.

[11] Farah Moety, Samer Lahoud, Bernard Cousin, and Kinda Khawam. Joint power-delay minimization in 4G wireless networks. In *Wireless Days (WD)*, 2014.

[12] 3GPP. Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Frequency (RF) system scenarios; v8.0.0 (Release 8). TR 36.942, 3rd Generation Partnership Project (3GPP), 2008-09.

[13] John E. Mitchell. *Handbook of Applied Optimization*, chapter Branch-and-Cut Algorithms for Combinatorial Optimization. Oxford University Press, 2002.

[14] 3GPP. Evolved Universal Terrestrial Radio Access (E-UTRA); Physical channels and modulation V8.4.0. TS 36.211, 3rd Generation Partnership Project (3GPP), 2008.

[15] 3GPP. Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) procedures in idle mode V8.2.0 (Release 8). TS 36.304, 3rd Generation Partnership Project (3GPP), 2005-08.

[16] D. Liu, L. Wang, Y. Chen, M. Elkashlan, K.-K. Wong, R. Schober, and L. Hanzo. User association in 5G networks: A survey and an outlook. *IEEE Communications Surveys Tutorials*, PP(99):1–1, 2016.

[17] IEEE Standard for Architectural Building Blocks Enabling Network-Device Distributed Decision Making for Optimized Radio Resource Usage in Heterogeneous Wireless Access Networks. *IEEE Std 1900.4-2009*, pages 1–130, 2009.