



HAL
open science

Queue-based model approximation for inter-cell coordination with service differentiation

Hadjer Touati, Hind Castel-Taleb, Badii Jouaber, Sara Akbarzadeh

► **To cite this version:**

Hadjer Touati, Hind Castel-Taleb, Badii Jouaber, Sara Akbarzadeh. Queue-based model approximation for inter-cell coordination with service differentiation. 24th Symposium on Computers and Communications (ISCC), IEEE, Jun 2019, Barcelona, Spain. pp.1-7, 10.1109/ISCC47284.2019.8969691 . hal-04492400

HAL Id: hal-04492400

<https://hal.science/hal-04492400v1>

Submitted on 6 Mar 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Queue-based model approximation for inter-cell coordination with service differentiation

Hadjer TOUATI¹, Hind CASTEL-TALEB¹, Badii JOUABER¹, and Sara AKBARZADEH²

¹SAMOVAR, Télécom SudParis, CNRS, 9 rue Charles Fourier, 91011 Evry Cedex, France

²DAVIDSON CONSULTING, 37 Rue Marcel Dassault, 92100 Boulogne-Billancourt, France

Abstract—This paper presents a queuing model to evaluate the performances of cooperation mechanisms in 4G/5G cellular networks. The main contribution is the proposal of an approximation model with a closed form formula allowing fast numerical evaluation of throughput and loss probabilities per service class, while considering traffic load, radio conditions and the available resources in each collaborating cell. The proposed queuing model is successfully applied to the enhanced Inter-Cell Interference Coordination mechanism with service differentiation in the context of Heterogeneous Cloud based Radio Access Networks. Numerical results are compared to Matlab simulations with realistic radio conditions and user distributions. Through it, decisions on resource allocation among the cooperating cells to achieve the optimal capacity of the system become faster and adaptive. In conclusion, the proposed queue-based model can be used in order to tune the front-haul links and the core network resources.

Index Terms—Queuing Model, Modeling, Performance, Inter-Cell, Cooperation, Approximation, Wireless, 5G, CRAN, HetNet, eICIC, Front-haul

I. INTRODUCTION

Future cellular networks are expected to offer higher spectral efficiency and capacity for mobile users, while being more flexible and adaptive for network providers. For this to happen, the 3GPP LTE-A standards introduced, since Rel-8, advanced interference mitigation techniques and cooperation mechanisms among neighbouring cells. These include the enhanced Inter-Cell Interference Coordination (eICIC) and the Coordinated Multi-Point (CoMP). In addition, two architectural concepts, namely the Heterogeneous Networks (HetNets) and Cloud (or Centralized) Radio Access Networks (CRAN) [1], are being considered and combined for future cellular networks. The first consists in mixing micro and macro cells in order to enhance radio conditions for edge users. The second proposes to split the base station functionalities into two parts according to different scenarios. For instance, only RF processing can be kept on “light” base stations denoted by RRH (Remote Radio Heads), while most of Base Band functions are processed on centralized and shared Base Band processing Units (BBUs). These architectural evolutions can ease implementing the

enhanced inter-cell interference mitigation and cooperation techniques by allowing centralized decisions based on global knowledge over a cluster of cells [1]. They can also lead to the reduction of both the OPEX and CAPEX for network providers.

To be efficiently deployed, this evolution requires new tools and approaches to analyze (and predict), in real time, the traffic requirements on the front-haul and the BBU. If not correctly sized, both the front-haul and the BBU pool may constitute a bottleneck that limits the RRHs from making full use of the available radio resources. In the following, we provide a traffic oriented model based on queuing theory and apply it to the eICIC/ABS (eICIC/ Almost Blank Sub-frame) mechanism. It allows deriving real-time performance of the eICIC/ABS according to traffic and radio parameters. It is an extension of the proposal in [2], where a Markov chain based model and an approximation model using the M/G/k/k queue were proposed. In this paper, we propose an approximation approach using a general closed form formula considering group arrivals and loss system with multiple services classes. The proposed model allows to jointly (i) determine the optimal eICIC/ABS configuration that maximizes the cell throughput and (ii) derive the expected throughput per service class. The proposed model is then used within a framework allowing to make fast decisions for the activation/deactivation of micro cells as well as for the distributions of sub-frames among macro and micro cells in case of activation.

The remainder of this paper is structured as follows: section II is dedicated to the analysis of the related work in the literature. In section III, we present the approximation model based on queuing theory and providing closed form formula. In section IV, we consider a case study and show that the proposed model provides accurate results compared to system level simulations. We then show how the proposed framework (model and algorithm) can be used for eICIC/ABS configuration in order to make fast decisions on the activation/deactivation of micro cells as well as for the

distributions of available resources among macro and micro cells. Then, in section V we present the proposed model-based optimization for cell coordination in the context of eICIC/ABS. Finally, section VI concludes the paper and presents some perspectives.

II. RELATED WORK

In the literature, many works investigated interference mitigation problems in HetNets using eICIC/ABS (Almost Blank Sub-frames). The main objective is to determine the optimal amount of radio resources that macro cells should delegate to their associated micro cells while ensuring maximum capacity. In [3], authors analyzed the interference using stochastic geometry to derive the required number of ABSs that guarantee a predetermined user QoS level. In [4], authors proposed two algorithms based on proportional fair metrics to dynamically adjust the number of ABS according to the instantaneous load conditions. In [5], an analytic framework, based on stochastic geometry, was proposed to allow micro cells avoiding interference to their neighboring macro cells. In [6], authors proposed a game theory based framework, in which the macro cell motivates small cells to perform ABS schemes by providing them a certain amount of monetary revenue. The problem is formulated as a one-leader multiple-follower Stackelberg game. In [7], another game theory based approach is proposed to jointly optimize the ABS scheme and small cell access control problem. A fuzzy logic system is adopted in [8] to compute the number of ABS. The inputs include UE (User Equipment) density, macro cell throughput and micro cell channel quality. In [9], the ABS configuration is formulated as a general-form consensus problem and then solved using the ADMM (Alternating Direction Method of Multiplier) approach. In [10], authors developed a mechanism to jointly adjust the parameters of ABS and the DRX (Discontinuous Reception) mechanism so as to improve LTE-A performance.

After the analysis of the related literature, it turns out that there is a lack of system level models allowing to dynamically analyze the eICIC performance. Such models can allow real-time tuning of the eICIC mechanism based on traffic load and users' radio conditions' measurements. It can help, for instance, to dynamically determine the appropriate distribution of available resources among macro and micro cells. In addition, most of the available studies mainly focus on maximum theoretical radio capacity and do not provide results at traffic and data flow levels.

III. QUEUING MODEL

We consider the M/G/k loss queuing system steady state distribution with batch arrivals [11]. In this system, groups of customers (a customer and a group of customers in this model refer respectively to a RBG and a client in our model) arrive at a service station in

accordance with a Poisson process of rate λ . The sizes of successive arriving groups are independent identically distributed random variables and independent of the arrival times.

Let $g_j, j = 1, 2, \dots, k$ be the group size probability distribution, and the corresponding tail probabilities is given by:

$$q_j = g_j + g_{j+1} + \dots (j = 1, 2, \dots, k) \quad (1)$$

The service is rendered by a number k of identical servers, each server being capable of serving one customer at a time. The system operates as follows: whenever a group of size j ($j = 1, 2, \dots, k$) finds on its arrival n ($n = 0, 1, \dots, k$) servers busy, then a number $\min(j, k-n)$ of its members, chosen at random, occupy an equal number of idle servers and start being served. The remaining customers, if any, are rejected by the system (i.e. blocked) and they do not return later. Hence no queue is allowed to be formed and the blocked customers are considered as lost to the system. A customer's service time (t_s) may depend on both the number and the service times of its group members accepted for service. However, customers who belong to different groups are supposed to have independent service times. In [11], D. Fakinos also considers the equilibrium behaviour of the M/G/k group-arrival group-departure loss system where the M/M/k group-arrival loss system is a particular case of this model. The equilibrium distribution (p_n), $n = 0, 1, \dots, k$, of the number of busy servers is given by:

$$p_n = \frac{\sum_{\substack{\nu_1, \dots, \nu_n \in \mathcal{N}_0 \\ \nu_1 + \dots + \nu_n = n}} \prod_{j=1}^n \frac{(q_j \rho / j)^{\nu_j}}{\nu_j!}}{\sum_{\substack{\nu_1, \dots, \nu_n \in \mathcal{N}_0 \\ \nu_1 + \dots + \nu_n \leq k}} \prod_{j=1}^k \frac{(q_j \rho / j)^{\nu_j}}{\nu_j!}} \quad (2)$$

where $\rho = \frac{\lambda}{\mu}$ is the traffic intensity of the system such that $\mu = \frac{1}{t_s}$ refers to the service rate.

From the probability distribution equation (2), we can derive X_j , the mean throughput for clients of size j . X_j is given by:

$$X_j = \lambda \times g_j \times \left(\sum_{l=0}^{k-j} j \times p_l + \sum_{l=k-j+1}^{k-1} (k-l) \times p_l \right) \quad (3)$$

Next, the proposed queuing model is applied for the analysis of the considered case study.

IV. eICIC MODEL AND PERFORMANCE ANALYSIS

In this section, we focus our studies on Downlink and consider a H-CRAN cluster composed of a central cell and its 6 first tier neighboring cells, numbered from 2 to 7 as shown in Fig 1. All these cells are macro cells. The cluster also includes a micro cell deployed within the central macro cell. We focus here our analysis on users within the central region, covered by both

the central macro and the associated micro cell and where the eICIC/ABS mechanism is used. Note that the interference from all the cluster cells is considered to determine users' radio conditions. We first consider the

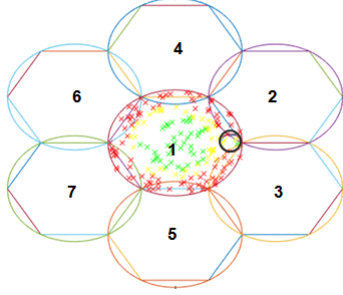


Fig. 1: H-CRAN cluster: 7 macro cells and 1 micro cell

general case where users are randomly and uniformly distributed over the central region. Algorithm 1 provides how to derive, for each user u , with a packet size of ($size(u)$) in bits, the required RBGs (Resource Block Groups) according to its radio conditions (for more details see [2]).

Algorithm 1: How to derive the required RBGs, for each user, in case of eICIC/ABS

Input: micro cell within a macro cell
Result: RBGs(u) for all users u

(a) Generation of users randomly and uniformly over the central macro cell;

for each user (u) do

(b) Determine the serving cell:
 $PRx_M(u) = Ptx_M(u) + G - PL_M(u)$
 $PRx_m(u) = Ptx_m(u) + G - PL_m(u)$
 $PRx_s(u) = \max(PRx_M(u), PRx_m(u))$

(c) Determine the SINR(u) with eICIC/ABS:
 $SINR(u) = 10 \log_{10} \left(\frac{PRx_s(u)}{\sum_{i=2}^7 PRx_i(u) + N_0} \right)$;

(d) Deduce the MCS(u) corresponding to SINR(u);

(e) Determine the number of required OFDM symbols: $N_{Symb}(u) = \frac{size(u)}{MCS(u)}$;

(f) Determine the corresponding number of RBs: $N_{Subch}(u) = \frac{N_{Symb}(u)}{7 \times 12}$;

(g) Determine the corresponding number of RBGs: $N_{RBGs}(u) = \frac{N_{Subch}(u)}{N_{RB}}$;

end

Parameters used in this algorithm are explained in TABLE I.

In the following, we denote by S the total number of available RBGs in a frame and N_{RB} the number of RBs within a RBG. These numbers depend on the considered bandwidth. The parameters that are used for evaluations are listed in TABLE II [13].

M, m, s, i	macro, micro, serving and interfering cells
PRx_M	received power from the macro (in watt)
Ptx_M	power transmission of the macro (in watt)
G	Gain of the antenna per Hz
N_0	Thermal noise
PL	pathloss model (in decibels): [12] $PL_M(u) = 128.1 + 37.6 \times \log_{10}(R)$ where R is in kilometers $PL_m(u) = 38 + 30 \times \log_{10}(R)$ where R is in meters

TABLE I: Users generation parameters

Parameter	Value
carrier frequency	2 GHz
bandwidth	20 MHz
macro cell radius	3500 m
micro cell radius	600 m
macro cell transmit power	40 dBm
micro cell transmit power	30 dBm
macro cell antenna gain	18 dBi
micro cell antenna gain	6 dBi
client size	1024 bits
n_{ABS}	3
S	500 RBGs/frame
N_{RB}	4 RBs
User location distribution	Uniform
Number of active users per frame	Poisson

TABLE II: Simulation parameters

A. Queuing model and parameters

Let λ_M (resp. λ_m) represents the client (i.e. data bursts) arrival rate to the macro cell (resp. micro cell). The total client arrival rate is then $\lambda = \lambda_M + \lambda_m$. The number of RBGs required per each client (according to its radio conditions) is determined through algorithm 1. This algorithm also defines the client class since we consider that a client requiring j RBGs belongs to class j , this means that this client has a batch size equal to j RBG and it requires j servers.

Let $\lambda_{M,j}$ and J_M (resp. $\lambda_{m,j}$, and J_m) be the arrival rate of class j and the number of classes in the macro cell (resp. micro cell). Therefore:

$$\lambda_M = \sum_{j=1}^{J_M} \lambda_{M,j} \text{ and } \lambda_m = \sum_{j=1}^{J_m} \lambda_{m,j} \quad (4)$$

Let S denote the number of servers in the queuing system. S is determined by the the number of RBG within an OFDM radio frame and depends on the selected radio bandwidth ($S=500$ servers for 20 MHz, 1000 servers for 40 MHz, ...). In the presence of a micro cell, these S servers are allocated among the macro and the micro cell following n_{ABS} : the number of ABS allocated to the micro cell.

Let S_M (resp. S_m) be the number of RBGs (and servers) within the macro cell (resp. within the micro cell). We have:

$$S_M = S - S_m \text{ and } S_m = S \times \frac{n_{ABS}}{10} \quad (5)$$

The behavior of eICIC technique can be modeled as a Markov chain [2]. However, since it considers batch arrivals and a relatively very high number of servers, it generates a very large number of transitions and its numerical resolution is costly in terms of time and resources. We then propose and discuss here an approximation model based on the M/G/k group arrival and loss system (described in section III).

The proposed M/G/k based approximation model is defined for both macro and micro cells.

For the macro cell, equation (2) is used with the following assumptions: $k = S_M$, $\rho = \rho_M = \frac{\lambda_M}{\mu_M}$, $\lambda = \lambda_M$, $\mu = \mu_M = 1$ (corresponding to one radio frame). The batch probabilities are obtained as follows: $g_j = \frac{\lambda_{M,j}}{\lambda}$, for $1 \leq j \leq J_M$. The mean throughput per class, $\bar{X}_{M,j}$, is then derived from equation (3). Similar reasoning is then done for the micro cell to determine the mean throughput $X_{m,j}$ by considering $k = S_m$, $\rho = \rho_m = \frac{\lambda_m}{\mu_m}$, $\lambda = \lambda_m$, $\mu = \mu_m = 1$.

B. Performance results and analysis

In this section, performance results are provided in terms of throughput and blocking probability per service class. Comparisons with system level simulations are conducted to estimate the accuracy of the proposed approximation. The used input parameters are derived from matlab simulations reproducing realistic behavior of the system (listed in TABLE II). All results are presented ensuring a 95% confidence intervals.

Two cluster scenarios were analyzed: with and without a micro cell. Two classes are considered requiring respectively 1 RBG and 2 RBG and such that $\lambda_{M,1} = 50\% \times \lambda_M$ and $\lambda_{M,2} = 50\% \times \lambda_M$.

The M/M/k group-arrival loss system parameters are derived as follows: $\lambda_M = \lambda_{M,1} + \lambda_{M,2}$, $g_1 = \frac{\lambda_{M,1}}{\lambda_M}$, $g_2 = \frac{\lambda_{M,2}}{\lambda_M}$, $q_1 = g_1 + g_2$, $q_2 = g_2$ and $\rho = \lambda_M$. Table III shows the accuracy of the closed form formula, by comparing the blocking probability (P_b) and the throughput (Th) per class when the macro is deployed without micro cell.

Results show that blocking probabilities per class, and consequently the corresponding throughput, obtained using the proposed M/M/k based approximations are quite close to those obtained by simulations.

While varying user arrivals from 50 to 350 per frame, the relative errors remain less than 1% for both classes. For higher rates, the relative error remains less than 6% for class 1 and 4% for class 2. Indeed, given that client departures from the M/G/k system are asynchronous, clients of class 2 (composed of two batches) leave the queuing system faster than in the real system as simulated, allowing the M/G/k group-arrival and loss system to serve more batches of class 1. Note that the relative error in terms of overall throughput remains below 1% for class 1 and below 3% for class 2.

λ	class 1		class 2	
	Simul (P_b)(Th)	Approx (P_b)(Th)	Simul (P_b)(Th)	Approx (P_b)(Th)
50	(0) (25)	(0) (25)	(0) (25)	(0) (25)
100	(0) (49)	(0) (49)	(0) (51)	(0) (51)
150	(0) (74)	(0) (74)	(0) (76)	(0) (76)
200	(0) (98)	(2.48E-08) (98)	(0) (102)	(4E-08) (102)
250	(6.72E-04) (123)	(2.66E-03) (123)	(0.0119) (128)	(0.0041) (128)
300	(6.72E-04) (147)	(2.66E-03) (147)	(0.0119) (152)	(0.0041) (152)
350	(0.0582) (164)	(0.0565) (164)	(0.1842) (166)	(0.0837) (161)
400	(0.1645) (165)	(0.1384) (171)	(0.2793) (168)	(0.1985) (162)
450	(0.2614) (164)	(0.2112) (175)	(0.3429) (168)	(0.2949) (160)

TABLE III: Comparison between Matlab simulations and the M/M/k group arrival in terms of blocking probability and throughput per class (macro without micro)

When the micro cell is deployed and the eICIC mechanism is active, the performances will obviously depend on the number of allocated ABSs and traffic and radio conditions. The optimal throughput can only be reached with optimal ABSs configuration.

To meet this objective, we propose in the following a model-based framework to ease decisions on the activation/deactivation of micro cells as well as for the distribution of sub-frames and clients among macro and micro cells when these latter are activated.

V. MODEL-BASED EICIC/ABS OPTIMIZATION

In this section, we propose a model based framework to dynamically decide about the activation/deactivation of micro cells as well as the optimal amount of radio resources macro cells should delegate to them. The main objective here is to maximize the overall system throughput.

When a micro cell is deployed, all users associated to it should experience almost no blocking (i.e. blocking probability $\leq \epsilon$). One can assume, without loss of generality, that radio conditions are always good for UEs associated to micro cells. Indeed, a UE is associated to a micro cell only if the latter allows enhancing its radio conditions. Otherwise the UE will remain associated to the macro cell. Therefore, one can assume that only one class of service is available in the micro cell, where all UEs require 1 RBG (i.e. $\lambda_m = \lambda_{m,1}$).

This optimization issue is formulated as an objective optimization problem. The decision variables are the following:

- λ_m : client arrival rate in the micro cell,
- λ_M : client arrival rate in the macro cell,
- S_m : the number of sub-frames to assign to the micro cell and

- $S_M = S - S_m$: the number of resources to assign to the macro cell.

Note that we deliberately set the client arrival rate to the micro cell as a variable since this rate (which is part of the overall arrival rate) should also be optimized together with the ABS configuration.

Since the radio conditions of edge clients are enhanced when micro is deployed, in the following, we consider the mean radio access bearer (RAB) or the MCS efficiency of the clients. The used RAB per class in the macro and the micro cells influences the optimization results. Let $RAB_M(j) = CR_M(j) \times \log_2(N_M(j))$ (resp. $RAB_m(1) = CR_m(1) \times \log_2(N_m(1))$) refers to the mean radio access bearer or the MCS efficiency of the users of class j in the macro (resp. of class 1 in the micro) where $CR_M(j)$ (resp. $CR_m(1)$) is the mean coding rate and $N_M(j)$ (resp. $N_m(1)$) is the mean number of constellation points of the users of class j in the macro (resp. of class 1 in the micro). These parameters per class are also extracted from simulations while real radio conditions are considered. Using the RAB will allow also converting the throughputs in bits/frame.

Our objective optimization problem is then expressed as follows:

$$\begin{aligned}
& \underset{S_m, S_M, \lambda_m, \lambda_M}{\text{maximize}} && f(S_m, S_M, \lambda_m, \lambda_M) \\
& \text{subject to} && S_m + S_M = S, \\
& && \lambda_m + \sum_{j=1}^{J_M} \lambda_{M,j} = \lambda, \\
& && \sum_{j=1}^{J_M} \lambda_{M,j} = \lambda_M, \\
& && S_M > 0, \\
& && S_m \geq 0, \\
& && p_m^b \leq \epsilon
\end{aligned} \tag{6}$$

where $f(S_m, S_M, \lambda_m, \lambda_M)$ represents the overall throughput expressed in terms of bits/frame and computed as follows:

$$\begin{aligned}
f(S_m, S_M, \lambda_m, \lambda_M) = & \sum_{j=1}^{J_M} \alpha_j \times j \times X_{M,j} \\
& + \beta \times X_{m,1}
\end{aligned} \tag{7}$$

so that:

$$\alpha_j = N_{RB} \times 7 \times 12 \times RAB_{M,j},$$

$$\beta = N_{RB} \times 7 \times 12 \times RAB_{m,1},$$

$$p_m^b = p_{S_m} \text{ and}$$

$X_{M,j}$ and $X_{m,1}$ are calculated by using equation(3).

Due to the combination of a continuous variable (such as λ_m) and discrete variables (such as S_m and S_M) and because the nonlinear expression in the objective function ($X_{M,j}$ for $j=1, \dots, J_M$ and $X_{m,1}$) and the

constraint (p_m^b), the maximization problem corresponds to a mixed integer non-linear program (MINLP). Non convexity as imposed by discrete variables or non convex nonlinear functions easily lead to problems that are both *NP*-hard in theory and computationally demanding in practice [14]. For detailed overview of MINLP solvers and the implemented algorithm, we refer to [15]. A simple way to solve this problem is to start by fixing S_m (and thereby S_M) and then, for each value we determine the optimal λ_m that maximizes the overall throughput. Finally, the optimal S_m (and S_M) and λ_m which offer the optimal overall throughput are retained.

Our problem corresponds now to a nonlinear program (NLP) due to the constraint and the objective function that are not linear. In Matlab, one manner of solving NLP is by using metaheuristic and a global optimization algorithm such as GlobalSearch and MultiStart that allow finding global minima. The following results are obtained using the GlobalSearch algorithm. It uses a Scatter Search mechanism for generating start points and then runs a local solver (fmincon) from multiple start points. Fmincon is a nonlinear programming solver that starts at initial point x_0 and attempts to find a minimizer x of the function described in the objective function subject to the constraints, x_0 can be a scalar, vector or matrix. The objective of Scatter Search is to maintain a set of diverse and high-quality candidate solutions. The principle of the approach is that useful information about the global optima is stored in a diverse and elite set of solutions (the reference set) and that recombining samples from the set can exploit this information. The strategy involves an iterative process, where a population of diverse and high-quality candidate solutions that are partitioned into subsets and linearly recombined to create weighted centroids of sample-based neighborhoods. The results of recombination are refined using an embedded heuristic and assessed in the context of the reference set as to whether or not they are retained [16].

Unfortunately, for our case study, these solvers take few days to provide results. In order to remedy this weakness, we propose to use an exhaustive search of the optimal values on a reduced set. We propose then algorithm 2 where:

(a) At first, the model-based optimization for eICIC/ABS determine X^* the overall throughput when the micro is not deployed according to equation (3).

(b) Furthermore, for each percentage of clients to be attached to the micro cell, the optimization model determines the minimum number of micro resources that achieves optimal throughput (determined by using equation (3)) with a service guarantee equal to $(1 - \epsilon)\%$ for micro clients. We note that i_{max} is the maximum percentage of clients to attach to the micro cell. The optimization model will therefore vary from i to i_{max} with a loop step noted 'step' (the smaller are the initial

value of i and the value of $step$, the better is the precision of the optimization model). Then, for each i , the model varies the number of ABS to be assigned to the micro cell from the minimum number of ABS ($n_{ABS} = 1$) to the maximum number of ABS noted ' $n_{ABS_{max}}$ '. Then, for each ABS, it determines the corresponding number of resources and keeps only those which guarantee a service of $(1-\epsilon)$ % for micro clients. After that, for each i , it selects the optimal number of micro resources which corresponds to the minimum that offers the maximum overall throughput.

After that in (c), it determines the unique i corresponding to the minimum of S_{m_i} maximizing the overall throughput.

In (d), it deduces the output parameters. Note that when $i < i_{max}$, the remaining clients of the micro cell ($i_{max} \times \lambda - i \times \lambda$) will be added to arrival rate of the macro in class J_M (λ_{M,J_M}), such that the total arrival rate λ remains constant.

The optimal throughput obtained in (d) corresponds to the scenario where micro is deployed. Thus in (e), it is compared to the overall throughput when micro is not deployed (determined in (a)).

Algorithm 2: How to determine the clients arrival rate and the number of servers in the micro cell to achieve the optimal overall throughput

Result: $S_m, S_M, \lambda_m, \lambda_M, X$

Input: $i, i_{max}, \lambda, J_M, \lambda_{M,j}$ for $j = 1, \dots, J_M, \epsilon, S, step$

(a) Compute the overall throughput in case of macro without micro (X^*);

(b) **while** $i \leq i_{max}$ **do**
 Determine the minimum S_{m_i} which maximizes the overall throughput, noted X_i , such that
 $p_m^b \leq \epsilon$;
 $i = i + step$;
end

(c) Determine i such that the corresponding $S_m = \min(S_{m_i})$ maximizes the overall throughput $X = \max(X_i)$;

(d) Deduce:

$$\lambda_m = i \times \lambda ;$$

$$\lambda_{M,J_M} = \lambda_{M,J_M} + (i_{max} \times \lambda - \lambda_m) ;$$

$$\lambda_M = \sum_{j=1}^{J_M-1} \lambda_{M,j} + \lambda_{M,J_M} ;$$

(e) Determine the optimal overall throughput $X = \max(X^*, X)$;

if $X = X^*$ **then**

$$S_m = 0 ; S_M = S ; \lambda_m = 0 ; \lambda_M = \lambda ;$$

end

According to the traffic load, the proposed based-model optimization for eICIC/ABS also allows deciding on the activation/ deactivation of the micro cell in order to achieve the optimal overall throughput.

In the following, we compare the results obtained by the algorithm 2 and the Global Search algorithm. We note in TABLE IV the optimal percentage of UEs accepted to be served by the micro, λ_m , and the optimal number of ABS, n_{ABS} , to assign to micro UEs, in order to reach the optimal overall throughput (X). Let's remember that the optimal number of servers assigned to the micro is $S_m = S \times \frac{n_{ABS}}{10}$ where $S = 500$ for a bandwidth of 20 MHz.

We consider a realistic distribution of clients over the macro and the micro with a uniform distribution, we have then: $\lambda_{M,1} = 50\% \times \lambda$ of the clients are at the center region (class 1) of the macro, $i_{max} \times \lambda = 3\% \times \lambda$ of the clients are at most in the center region (class 1) of the micro (with good radio conditions), $\lambda_{M,2} = 47\% \times \lambda$ of the clients are at the edge region (class 2) of the macro (with bad radio conditions).

We limit the search space of the optimal number of servers by limiting the number of possible ABS to 5 (half of the available ABS in a frame). We assume that $i = 1\%$, $step = 1\%$ and $\epsilon = 0.02$ the purpose of this condition is to ensure that micro UEs are served at 98%. Note that realistic values of user's MCS efficiency per class (extracted from Matlab simulations) are used where: $RAB_M(1) = 3.7746$, $RAB_M(2) = 0.86$ and $RAB_m(1) = 3.334$.

Clients arrival rate (λ)	Arrival rate [kbits/frame]	Optimal n_{ABS}	Optimal λ_m	Optimal X
50	46	1	$3\% \times \lambda$	46
100	92	1	$3\% \times \lambda$	92
200	183	1	$3\% \times \lambda$	183
300	275	1	$3\% \times \lambda$	270
400	361	0	$0\% \times \lambda$	305
500	451	0	$0\% \times \lambda$	315
600	541	0	$0\% \times \lambda$	322
700	631	0	$0\% \times \lambda$	327
800	734	1	$3\% \times \lambda$	333
900	826	1	$3\% \times \lambda$	340
1000	917	1	$3\% \times \lambda$	346
1200	1101	1	$3\% \times \lambda$	357

TABLE IV: Optimal n_{ABS} , optimal λ_m and optimal overall throughput in kbits/frame according to the clients arrival rates obtained by Global Search algorithm and algorithm 2

The same optimal values are obtained by the Global Search algorithm and algorithm 2, therefore, the results are only shown once in TABLE IV. Note that algorithm 2 offers the results more quickly. For each arrival rate (or user density in kbits/frame), an optimal number of n_{ABS} and an optimal number of micro UEs can be selected in order to maximize the overall throughput in terms of kbits/frame (as explained in the objective optimization problem). When the client arrival rate varies between 46 and 275 kbits/frame, the macro and the micro are lightly loaded. The algorithms propose to activate the micro cell and to serve all the micro UEs. Hence, the

macro allocates the necessary number of ABS to the micro since there are enough resources in the macro to serve their own UEs. Note that 1 ABS for the micro is enough to serve their clients and $p_m^b \leq 0.02$, due to the small percentage of clients in the micro (3%).

When the client arrival rate is greater than 361 and less than 631 kbits/frame, the optimization models propose to not deploy the micro. In fact, for these clients arrival rates the macro cell is loaded but the micro is not. The allocation of 1 ABS to the micro will penalize the macro UEs which decrease the overall throughput.

When the client arrival rate is greater than 734 kbits/frame, the macro is very loaded and the micro is moderately loaded. The deployment of the micro cell will allow to enhance the overall throughput however the optimization algorithms propose to delegate to the micro the minimum of ABS that it needs to reach the optimal overall throughput and by assigning the maximum number of clients such that $p_m^b \leq 0.02$.

VI. CONCLUSION

In this paper, we propose a queue-based model approximation with service differentiation for cell coordination. The accuracy of the proposed model is validated and compared to system level simulations. The relevance of the model is that the input parameters are based on real user distributions and radio conditions (measured from real systems or extracted from simulations). The model is applied to the eICIC/ABS mechanism and numerical results are presented. The proposed queuing model is then used to develop an optimization framework able to adapt the eICIC/ABS parameters according to user load and radio conditions. As a future works, we plan to compare our results with NS3 simulations using general arrival distributions. Furthermore, we plan to study the distribution of resources over coordinated cells within a cluster with multiple tiers. Such a study will allow as to define the traffic load in order to dimension the front-haul links and the BBU pool in H-CRAN. In addition, when considering latency and jitter requirements for various traffic flows, front-haul links need to be tuned leveraging statistical multiplexing gains, infrastructure reuse, and ultimately, cost reduction. To this end, we plan to combine our model with the one proposed in [17], where authors proposed a queue based model for the front-haul links.

REFERENCES

- [1] M. Peng, Y. Li, Z. Zhao, and C. Wang, "System architecture and key technologies for 5g heterogeneous cloud radio access networks," *IEEE network*, vol. 29, no. 2, pp. 6–14, 2015.
- [2] H. Touati, H. Castel-Taleb, B. Jouaber, and S. Akbarzadeh, "Simulation, modeling and analysis of the eicic/abs in h-crans," in *Performance Evaluation and Modeling in Wired and Wireless Networks (PEMWN)*, *International Conference on*. IEEE, 2018, pp. 1–6.
- [3] M. Cierny, H. Wang, R. Wichman, Z. Ding, and C. Wijting, "On number of almost blank subframes in heterogeneous cellular networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 10, pp. 5061–5073, 2013.
- [4] B. Soret and K. I. Pedersen, "Centralized and distributed solutions for fast muting adaptation in lte-advanced hetnets," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 1, pp. 147–158, 2015.
- [5] T. Zhou, Y. Chen, C. Jiang, and K. R. Liu, "Pricing game for time mute in femto-macro coexistent networks," *IEEE Transactions on Wireless Communications*, vol. 14, no. 4, pp. 2118–2130, 2015.
- [6] L. Li, Z. Zhou, Y. Hu, T. Jiang, and M. Wei, "Pricing framework for almost blank subframe scheme in two-tier heterogeneous networks," *International Journal of Communication Systems*, vol. 31, no. 3, p. e3454, 2018.
- [7] L. Li, Z. Zhou, S. Sun, and M. Wei, "Distributed optimization of enhanced intercell interference coordination and resource allocation in heterogeneous networks," *International Journal of Communication Systems*, p. e3915.
- [8] A. Daeinabi, K. Sandrasegaran, and P. Ghosal, "An enhanced intercell interference coordination scheme using fuzzy logic controller in lte-advanced heterogeneous networks," in *2014 International Symposium on Wireless Personal Multimedia Communications (WPMC)*. IEEE, 2014, pp. 520–525.
- [9] H. Zhou, Y. Ji, X. Wang, and S. Yamada, "eicic configuration algorithm with service scalability in heterogeneous cellular networks," *IEEE/ACM Transactions on Networking (TON)*, vol. 25, no. 1, pp. 520–535, 2017.
- [10] Y.-C. Wang and C.-C. Huang, "Efficient management of interference and power by jointly configuring abs and drx in lte-a hetnets," *Computer Networks*, vol. 150, pp. 15–27, 2019.
- [11] D. Fakinos, "On the m/g/k group-arrival loss system," *European journal of operational research*, vol. 44, no. 1, pp. 75–83, 1990.
- [12] 3GPP TR 36.931 version 14.0.0 Release 14, "Radio frequency (RF) requirements for lte pico nodeB."
- [13] 3GPP TS 36.213 version 9.1.0 Release 9, "Physical layer procedures."
- [14] S. Vigerske and A. Gleixner, "Scip: Global optimization of mixed-integer nonlinear programs in a branch-and-cut framework," *Optimization Methods and Software*, vol. 33, no. 3, pp. 563–593, 2018.
- [15] J. Kronqvist, D. E. Bernal, A. Lundell, and I. E. Grossmann, "A review and comparison of solvers for convex minlp," *Optimization and Engineering*, pp. 1–59, 2018.
- [16] F. Glover, "A template for scatter search and path relinking," *Lecture notes in computer science*, vol. 1363, pp. 13–54, 1998.
- [17] G. O. Pérez, J. A. Hernández, and D. Larrabeiti, "Fronthaul network modeling and dimensioning meeting ultra-low latency requirements for 5g," *Journal of Optical Communications and Networking*, vol. 10, no. 6, pp. 573–581, 2018.