

# Analytical Model for the UE Blocking Probability in an OFDMA Cell providing GBR Slices

Oscar Adamuz-Hinojosa<sup>\*†</sup>, Pablo Ameigeiras<sup>\*†</sup>, Pablo Muñoz<sup>\*†</sup>, Juan M. Lopez-Soler<sup>\*†</sup>

<sup>\*</sup>Research Center on Information and Communication Technologies, University of Granada.

<sup>†</sup>Department of Signal Theory, Telematics and Communications, University of Granada.

Email: {oadamuz, pameigeiras, pabloml, juanma}@ugr.es<sup>\*†</sup>

**Abstract**—When a network operator designs strategies for planning and operating Guaranteed Bit Rate (GBR) slices, there are inherent issues such as the under(over)-provisioning of radio resources. To avoid them, modeling the User Equipment (UE) blocking probability in each cell is key. This task is challenging due to the total required bandwidth depends on the channel quality of each UE and the spatio-temporal variations in the number of UE sessions. Under this context, we propose an analytical model to evaluate the UE blocking probability in an Orthogonal Frequency Division Multiple Access (OFDMA) cell. The main novelty of our model is the adoption of a multi-dimensional Erlang-B system which meets the reversibility property. This means our model is insensitive to the holding time distribution for the UE session. In addition, this property reduces the computational complexity of our model due to the solution for the state transition probabilities has product form. The provided results show that our model exhibits an estimation error for the UE blocking probability below 3.5%.

**Index Terms**—Blocking probability, OFDMA, GBR, Erlang-B

## I. INTRODUCTION

Nowadays, the industry digitalization has boosted a wide variety of unprecedented services with stringent requirements. To economically provide them over a common infrastructure, network slicing has emerged as a solution [1]. Implemented as slices, most of these services are envisioned to rely on data transmissions with a strict Guaranteed Bit Rate (GBR) for each User Equipment (UE). When a network operator designs planning and operational strategies for GBR slices, it must consider the specific bandwidth consumption of each active UE per slice as well as the spatio-temporal variations in the number of UE sessions. Designing these strategies is challenging due to the bandwidth consumption of each UE is conditioned to its channel quality. With the aim of maintaining the Block Error Rate (BLER) for the UE's data below a certain threshold, the cells adopt Link Adaptation (LA). This technique enables each cell to adapt the UEs' Modulation and Coding Scheme (MCS) according to the experienced channel effects (i.e., path loss, shadowing, fast fading, inter-cell interference) [2].

This work is partially supported by the H2020 research and innovation project 5G-CLARITY (Grant No. 871428); the Spanish Ministry of Economy and Competitiveness, the European Regional Development Fund (Project PID2019-108713RB-C53); and the Spanish Ministry of Education, Culture and Sport (FPU Grant 17/01844).

To avoid inherent issues such as the under(over)-provisioning of radio resources, modeling the UE blocking probability in each cell is crucial for designing planning and operational strategies for GBR slices. Thereby, the network operator could decide the required number of cells, including their bandwidths, to deploy/scale GBR slices while the UE blocking probability is below a certain threshold.

Valuable models for evaluating the UE blocking probability have already been proposed in the literature (reported in Section II). Most of them rely on Markov chains and queue theory. However, they are not appropriate for GBR slices due to they consider a variable rate per each UE. Additionally, some of these models are only valid for UE session durations following an exponential distribution, thus they cannot model the traffic behavior in real scenarios [3].

In this article, we focus on modeling the radio resource consumption of a GBR slice. Specifically, we have proposed an analytical model for assessing the UE blocking probability in a GBR slice for an Orthogonal Frequency-Division Multiple Access (OFDMA) cell under Poisson session arrivals. Using our model, the network operator can decide the number of radio resources required by a cell to provide a GBR slice while the UE blocking probability is below a given threshold. The main novelty of our model is the employment of a Multi-dimensional Erlang-B system which meets the reversibility property. It means our model allows the adoption of an arbitrary distribution for the UE session duration. Additionally, this property involves the solution for the state probabilities has product form, thus it eases their computation. Another innovation is the consideration of the average Signal-to-Interference-plus-Noise Ratio (SINR) for each UE. This allows a more precise characterization of the UEs' channel quality within the cell. The provided results show that our model exhibits an estimation error for the UE blocking probability below 3.5%.

The article is organized as follow. Section II summarizes the relevant literature. Section III presents the system model. In Section IV, we present the proposed Multi-dimensional Erlang-B model. This model is validated in Section V. Finally, Section VI draws the main conclusions and the future work.

## II. RELATED WORK

The existing literature for modeling the UE blocking probability in a cell is vast. In [4], the authors model a Code

Division Multiple Access (CDMA)-High Data Rate (HDR) cell with a multi-class processor sharing queue. Since the processor sharing discipline is insensitive to the holding time distribution, arbitrary distributions can be adopted for the UE session duration. This discipline also forces an equal distribution of radio resources among the UEs, thus this model properly captures the behavior of Variable Bit Rate (VBR) services. In [5], this work was extended by including intra-cell UE mobility. However this improvement involves losing the insensitivity property, thus only the exponential distribution is valid for the UE session duration.

In both works, the authors have considered concentric rings to model the channel quality distribution within the cell. This approximation easily enables the model to capture the behavior of LA techniques. For this reason, others authors have adopted this approach. In [6]–[8], the authors also focus on scenarios with intra-cell UE mobility but considering other medium access techniques such as Wideband CDMA (WCDMA) and OFDMA. Other works concentrate on non-3GPP access technologies. For instance, the authors of [9] define a model for WiMAX cells. This model is then used by a Quality of Service (QoS)-oriented resource allocation strategy for streaming flows that require a constant bit rate. Additionally, other authors consider services beyond mobile broadband. For example, the authors of [10], [11] adapt their models to IPTV services.

Despite these works present valuable contributions, the consideration of concentric rings limits the accuracy for modeling the distribution of the channel quality. For instance, two UEs located to the same distance from an access node could not perceive the same channel quality. The reason is there could be different obstacles and geographical features between each UE and the access node, involving a different impact of the channel effects such as shadowing or fast fading.

In an attempt to improve the model for the channel quality distribution, the authors of [12] consider a combination of indicators such as the RSRP, RSRQ, RSSI and SINR for each UE. Then, they use these indicators in a Markov chain which models the operation of a cell with intra-cell UE mobility. Notwithstanding, the authors assume the reduction of the UE data rate when the total required bandwidth exceed the available bandwidth in the cell, thus this model is more appropriate for VBR services.

### III. SYSTEM MODEL

In this work, we focus on the downlink operation of one OFDMA cell. It provides a GBR service to their UEs, which dynamically request and release data sessions. This cell also supports LA, thus it must consider the channel quality perceived by each UE to allocate them radio resources. Based on this scenario, we first present the model for the cell. Then, we define the characteristics of the offered traffic. Finally, we describe the model for the radio resources.

#### A. Cell Model

To measure the channel quality within the cell, we adopt the SINR. This indicator depends on the radio environment and

varies over time mainly due to (a) the path loss propagation, (b) shadowing, (c) fast fading and (d) inter-cell interference. In this work, we assume UEs have reduced mobility within the cell (e.g., semi-static people in live events such as sport events or concerts, IoT sensors, equipment for industry 4.0), thus (a)-(b) remain constant throughout the session duration while (c)-(d) vary over time. Let us define  $\gamma_{u,n}^{(t)}$  as the instantaneous SINR measured by the UE  $u$  in the Resource Block (RB)  $n$  (see section III-C) from the cell  $c$ . This parameter is provided by Eq. 1, where  $P_c^{RX}$  denotes the received power from the access node. This power results from the transmitted power less the attenuation suffered by the channel effects.  $L_j$  is the cell load factor and  $\alpha_{j,n}$  is a function that takes the value 1 when the RB  $n$  is allocated to the neighbor cell  $j$  and the value 0 otherwise [1]. Note that the value for  $\alpha_{j,n}$  will depend on the radio resource allocation algorithm implemented in each neighbor cell. Finally  $P_N$  is the noise power measured in one RB.

$$\gamma_{u,n}^{(t)} = \frac{P_c^{RX}}{\sum_{j \in \mathcal{C} \setminus \{c\}} L_j \alpha_{j,n} P_j^{RX} + P_N} \quad (1)$$

In this work, we focus on the average SINR for each UE  $\bar{\gamma}_u$ . To obtain this parameter, we average  $\gamma_{u,n}^{(t)}$  over all the RBs consumed by the UE  $u$  (i.e.,  $\forall n \in \mathcal{N}_{RB}^{u,(t)}$ ), and the time period  $T_{ff}$  as Eq. 2 shows. This period is a time window over the fast fading is distinguishable. We also assume if  $\bar{\gamma}_u$  is measured several times throughout the UE session, its remains constant. Note that the operator  $|\cdot|$  denotes the cardinality of a set (i.e., the number of elements).

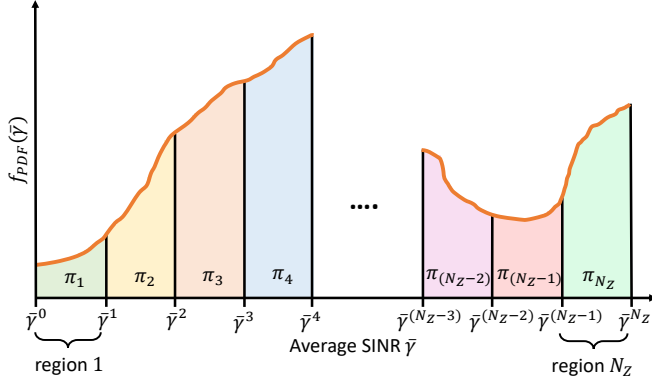
$$\bar{\gamma}_u = \frac{1}{T_{ff}} \int_{\tau}^{\tau+T_{ff}} \frac{1}{|\mathcal{N}_{RB}^{u,(t)}|} \sum_{n \in \mathcal{N}_{RB}^{u,(t)}} \gamma_{u,n}^{(t)} dt \quad (2)$$

Considering  $\bar{\gamma}_u$  is measured for a considerable amount of UEs, we can derive the Probability Density Function (PDF) for the average SINR  $f_{PDF}(\bar{\gamma})$  to model the channel quality within the cell. Since this feature could take a huge number of values, we split them into  $N_Z$  regions to make it treatable. Depicted in Fig. 1, each region  $z$  is defined as the set of values for the average SINR such as  $\bar{\gamma} \in [\bar{\gamma}^{(z-1)}, \bar{\gamma}^z]$ . For simplicity, we assume the session of an active UE takes place in one of these  $N_Z$  regions with probability  $\pi_z$ , which is provided by Eq. 3. Note that  $\sum_{z=1}^{N_Z} \pi_z = 1$ .

$$\pi_z = \int_{\bar{\gamma}^{(z-1)}}^{\bar{\gamma}^z} f_{PDF}(\bar{\gamma}) d\bar{\gamma} \quad (3)$$

Finally, we can also derive the average data rate per bandwidth unit  $\overline{SE}_z$  (i.e., spectral efficiency) for each region  $z$  by using Eq. 4. The function  $f_{SINR \rightarrow SE}(\bar{\gamma})$  maps the SINR to the spectral efficiency under the assumption each UE achieves all the required radio resources. This function depends on the LA technique employed in the cell.

$$\overline{SE}_z = \int_{\bar{\gamma}^{(z-1)}}^{\bar{\gamma}^z} f_{SINR \rightarrow SE}(\bar{\gamma}) f_{PDF}(\bar{\gamma}) d\bar{\gamma} \quad (4)$$


 Fig. 1. Splitting  $f_{PDF}(\bar{\gamma})$  into  $N_Z$  regions

### B. Traffic Model

To model the traffic demands within a cell, we consider the statistical distributions and the average values for the arrival rate of UE sessions and the session duration.

For the arrival rate, we assume an average of  $\lambda$  UE session requests per unit time following a Poisson distribution. Since a Poisson process can be split into  $N_Z$  independent process [13], we can also express the average arrival rate for each region as  $\lambda_z = \lambda\pi_z$ . Note that  $\lambda = \sum_{z=1}^{N_Z} \lambda_z$ .

With respect to the session duration  $t_{s,u}$  for each UE  $u$ , we assume a random variable extracted from an arbitrary distribution. Additionally, we define  $\mu = 1/E[t_{s,u}]$  as the average rate for releasing UE sessions per unit time. Note that the release rate is independent from the region  $z$ .

### C. Radio Resource Model

We assume an OFDMA cell with a total bandwidth  $W$ . This bandwidth is divided into  $N$  OFDM sub-carriers. In turn, these sub-carriers are grouped in groups of  $N_{SC}$  sub-carriers. Each group defines a RB, which is the smallest unit of resources that can be allocated to a UE. The number of available RBs on average during a slot is given by Eq. 5. The parameter  $\Delta f$  is the bandwidth between sub-carriers whereas  $OH$  denotes the overhead factor due to control plane data.

$$N_{RB}^{slot} = \left\lfloor \frac{W}{N_{SC}\Delta f}(1 - OH) \right\rfloor \quad (5)$$

Assuming all the UEs require an average data rate equal to the service GBR  $D_{GBR}$ , we need to compute the average number of RBs for each UE  $u$  in a time slot. Since we assume each UE is born in a specific region  $z$ , there exist only  $N_Z$  values for the average number of RBs required by a single UE within the cell. These values are given by Eq. 6 and must satisfy Eq. 7 for any UE which is born in a region  $z$  (i.e.,  $U^z$ ). In Eq. 7,  $L_{n,u}$  denotes the specific amount of RBs allocated to the UE  $u$  in each time slot, whereas  $N_{slots}^u$  denotes the number of time slots during the session duration  $t_{s,u}$ . We assume  $L_{n,u}$  is determined by a scheduler which aims to meet the GBR requirements of each UE.

$$N_{RB,z}^{slot} = \left\lceil \frac{D_{GBR}}{\overline{SE}_z N_{SC} \Delta f} \right\rceil \quad (6)$$

$$N_{RB,z}^{slot} = \frac{1}{N_{slots}^u} \sum_{n=1}^{N_{slots}^u} L_{n,u} \quad \forall u \in U^z \quad (7)$$

## IV. CAPACITY MODEL OF AN OFDMA CELL

This section explains the proposed model for an OFDMA cell, including the methodology used for deriving the UE blocking probability, the average RB utilization, and the cell capacity.

### A. Multi-dimensional Erlang-B model

Let us consider a cell where the values of  $f_{PDF}(\bar{\gamma})$  are grouped into  $N_Z$  regions. To model this system we employ a multi-dimensional Erlang-B system. In this model, we assume each UE session takes place into one region  $z$ , defined by the tuple  $(\lambda_z, \mu)$ . The offered traffic intensity in each region becomes  $\rho_z = \lambda_z/\mu$ , and the total offered traffic intensity is  $\rho = \sum_{z=1}^{N_Z} \rho_z$ .

Let  $s = (U_1, U_2, \dots, U_{N_Z})$  denotes the state of the system, where  $U_z$  is the number of active UEs in the region  $z$ . To define the set of feasible states, we take into account (a) an active UE in the region  $z$  consumes  $N_{RB,z}^{slot}$  RBs to meet its requirements; and (b) the available RBs in the cell are limited by  $N_{RB}^{slot}$ . These statements are gathered by Eq. 8, which provides the necessary condition to define a feasible state.

$$N_{RB}^{slot} - \sum_{z=1}^{N_Z} U_z N_{RB,z}^{slot} \geq 0 \quad \forall s \quad (8)$$

Using this equation, we can build the state transition diagram as Fig. 2 shows. Note that for understandability purposes, the represented diagram only shows two dimensions. To define the upper bounds in each dimension, we use Eq. 9. In this equation,  $U_{z|U_y}^{cmax}$  denotes the maximum number of UEs in the region  $z$  conditioned to the number of UEs in the remaining regions (e.g., red states for region 1, and green states for region 2). If the remaining regions have 0 UEs, we can define the absolute maximum number of UEs in region  $z$  as  $U_z^{amax} = \left\lfloor \frac{N_{RB}^{slot}}{N_{RB,z}^{slot}} \right\rfloor$ .

$$U_{z|U_y}^{cmax} = \left\lfloor \frac{N_{RB}^{slot} - \sum_{y \in Z \setminus \{z\}} U_y N_{RB,y}^{slot}}{N_{RB,z}^{slot}} \right\rfloor \quad (9)$$

To clarify how the state transition diagram is built, Fig. 3 depicts a specific realization of the two-dimensional Erlang-B system presented in Fig. 2. In this example, each UE consumes  $D_{GBR} = 2$  Mbps. Considering the average spectral efficiencies derived by Eq. 4 for each region are  $\overline{SE}_1 = 2.778$  bps/Hz and  $\overline{SE}_2 = 1.389$  bps/Hz, the required amount of RBs in each region are  $N_{RB,1}^{slot} = 4$  and  $N_{RB,2}^{slot} = 8$  (i.e., using Eq. 6 and considering  $\Delta f = 15$  KHz and  $N_{SC} = 12$  sub-carriers). If we assume the cell has available  $N_{RB}^{slot} = 20$  RBs, we can check how Eq. 8 is met for all the states presented in Fig. 3.

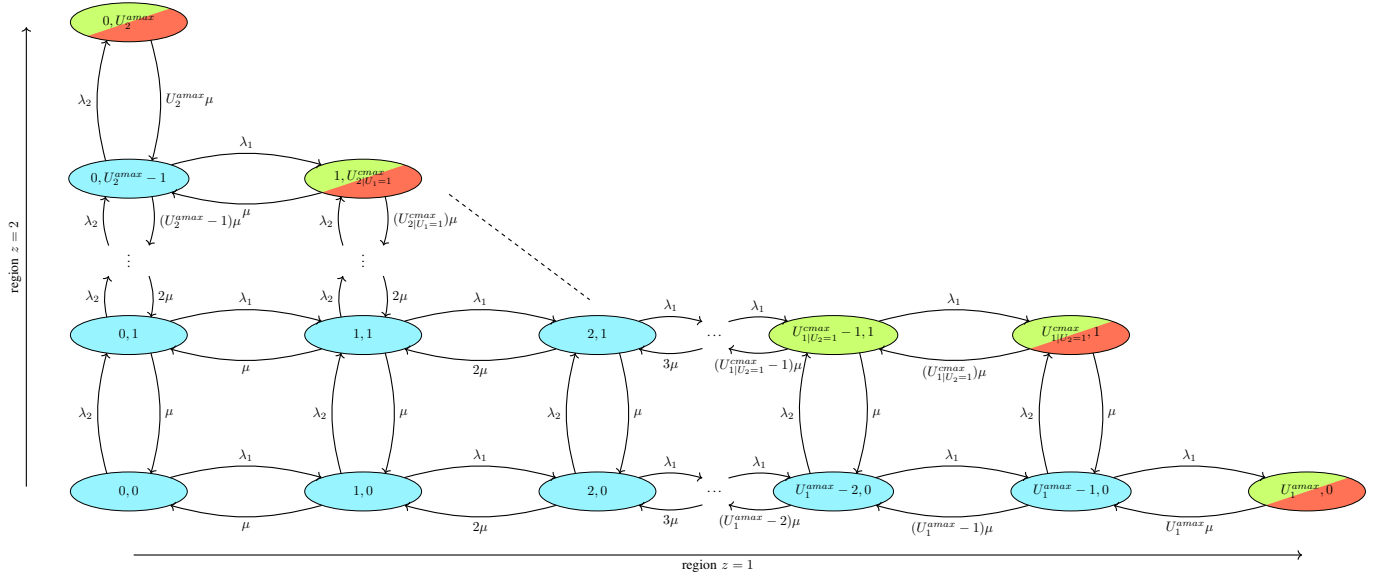


Fig. 2. State transition diagram for a two-dimensional Erlang-B system. Note that red and green states correspond to  $U_1 = U_{1|U_2}^{cmax}$  and  $U_2 = U_{2|U_1}^{cmax}$ , respectively.

Furthermore, the absolute maximum number of UEs in each region (i.e., when the number of UEs in the remaining regions is 0) is  $U_1^{amax} = 5$  UEs and  $U_2^{amax} = 2$  UEs, respectively. Finally, focusing on the upper bounds in each dimension (i.e., red and/or green states), we can verify how Eq. 9 is met. For instance, considering region  $z = 1$ , the upper bounds (i.e., red states) are (a)  $U_{1|U_2=0}^{cmax} = 5$  when  $U_2 = 0$ ; (b)  $U_{1|U_2=1}^{cmax} = 3$  when  $U_2 = 1$ ; and (c)  $U_{1|U_2=2}^{cmax} = 1$  when  $U_2 = 2$ .

The resulting multi-dimensional Erlang-B system corresponds to a reversible Markov process (see proof in appendix A). This implies the proposed model is insensitive to the distribution of the UE session duration, which means the state probabilities depend only upon the mean service time [13]. Furthermore, the solution for the state probabilities has product form as Eq. 10 shows, where  $p(U_z)$  is the one-dimensional truncated Poisson distribution for traffic stream in region  $z$  and  $K$  is a normalization constant.

$$p(U_1, U_2, \dots, U_{N_Z}) = K \cdot p(U_1) \cdot p(U_2) \cdot \dots \cdot p(U_{N_Z}) \\ = K \cdot \prod_{z=1}^{N_Z} \frac{\rho_z^{U_z}}{U_z!} \quad (10)$$

To obtain the state probabilities, we need to derive  $K$ . This constant can be computed by summing all the state probabilities and equating the resulting expression to 1, i.e.,  $\sum_{\forall s} p(U_1, U_2, \dots, U_{N_Z}) = 1$ . To that end, we recursively cover all the feasible states by using Eq. 11.

$$K^{-1} = \sum_{U_{N_Z}=0}^{U_{N_Z}^{amax}} \sum_{U_{N_Z-1}=0}^{U_{N_Z-1}^{uplim}} \dots \\ \sum_{U_1=0}^{U_{1|U_2, \dots, U_{N_Z}}^{uplim}} \left( \prod_{z=1}^{N_Z} \frac{\rho_z^{U_z}}{U_z!} \right) \quad (11)$$

In this equation, there is a summation per dimension with the aim of covering all the feasible states. The iterator of each summation is the number of UEs per a specific dimension, and it is bounded by the upper limit given in Eq. 12. This limit is an extension of  $U_{z|U_y}^{cmax}$  (see Eq. 9) which restricts its inter-dimensional dependence to those regions above  $z$ , i.e., the iterators of the outer summations. In this way, when the iterators of the outer summations increase, these summations will not cover those probabilities previously covered by the previous iterators. Considering the example provided in Fig. 3, this equation will iteratively cover the states of each row from the bottom to the top.

$$U_{z|U_{N_Z}, U_{N_Z-1}, \dots, U_x}^{uplim} = \left\lfloor \frac{N_{RB}^{slot} - \sum_{y=z+1}^{N_Z} U_y N_{RB,y}^{slot}}{N_{RB,z}^{slot}} \right\rfloor \quad \forall x > z \quad (12)$$

## B. UE Blocking Probability

Assuming a new UE session is born in region  $z$ , it will be blocked if there not exists a transition from the current state  $s^{(t)} = (U_1, U_2, \dots, U_z, \dots, U_{N_Z})$  to  $s^{(t+1)} = (U_1, U_2, \dots, U_z + 1, \dots, U_{N_Z})$ . This happens when  $U_z + 1 > U_{z|U_y}^{cmax}$ . In this way, the blocking states for each dimension

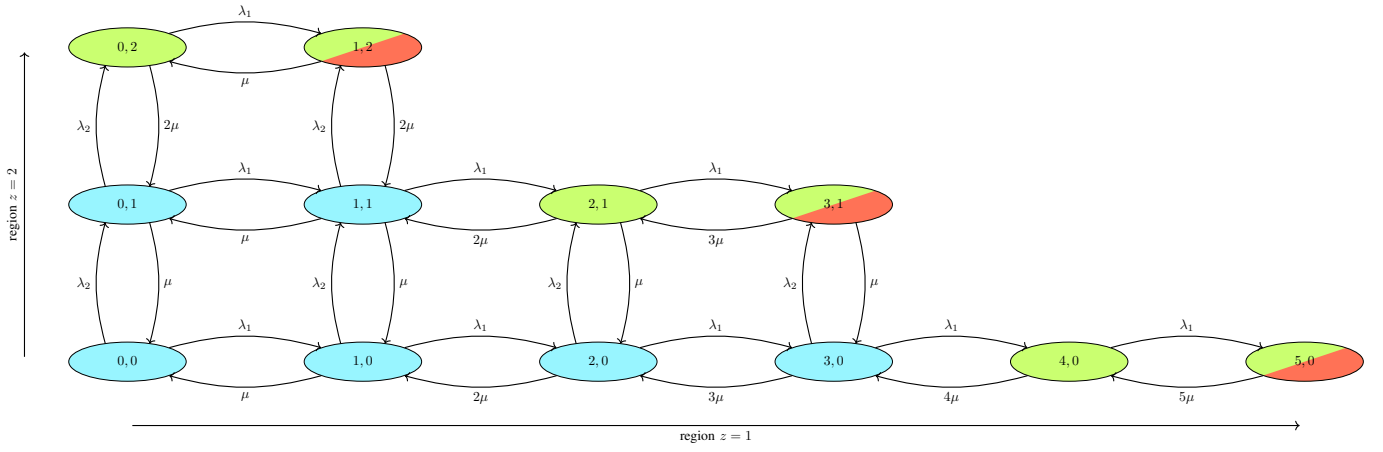


Fig. 3. Specific realization of the state transition diagram for a two-dimensional Erlang-B system.

are delimited by Eq. 9 (e.g., red states for region 1 in Fig. 3). If we iteratively cover the blocking states for each dimension, we can compute the UE blocking probability  $B_z$  conditioned to the region  $z$  where the new UE session is born by Eq. 13.

$$B_z = \sum_{U_{N_z}=0}^{U_{N_z}^{max}} \sum_{U_{N_z-1}=0}^{U_{N_z-1}^{bplim}} \dots \sum_{U_1=0}^{U_1^{bplim}} p(U_1, U_2, \dots, U_z^{cmx}, \dots, U_{N_z}) \quad (13)$$

This expression is composed by  $N_z - 1$  recursive summations, one per region excluding  $z$ . In turn, each summation is delimited by  $U_{k|z, U_{N_z}, \dots, U_x}^{bplim}$ , given by Eq. 14. The aim of this upper bound is similar to Eq. 12 with the difference the region  $z$  is not considered (i.e.,  $U_z$  must be forced to  $U_z^{cmx}$ ). Thereby, the summations only cover the blocking states in each iteration. The variable  $k$  denotes the specific region that a summation covers.

$$U_{k|z, U_{N_z}, \dots, U_x}^{bplim} = \left\lfloor \frac{N_{RB}^{slot} - \sum_{y>k, y \in \mathcal{Z} \setminus \{z\}} U_y N_{RB,y}^{slot}}{N_{RB,k}^{slot}} \right\rfloor \quad \forall x > k \quad (14)$$

Finally, the UE blocking probability in the cell is computed as the sum of the conditional blocking probabilities weighted by the probability of a UE session is born in each region (see Eq. 15).

$$B = \sum_{z=1}^{N_z} \pi_z B_z \quad (15)$$

### C. Mean number of consumed RBs and Cell Capacity

Another key parameters derived by our model are the mean number of RBs consumed in a cell  $\bar{N}_{RB}$ , and the cell capacity  $D_c$ .

The mean number of RBs can be computed as the average number of UEs  $\bar{U}_z$  in each region  $z$  multiplied by the consumed RBs (see Eq. 16). In turn, we compute  $\bar{U}_z$  using Little's theorem [13], i.e.,  $\bar{U}_z = \lambda_z / \mu (1 - B_z) = \rho_z (1 - B_z)$ .

$$\bar{N}_{RB} = \sum_{z=1}^{N_z} \bar{U}_z N_{RB,z}^{slot} \quad (16)$$

The cell capacity is provided by Eq. 17. It is derived as the product of the mean number of UEs in each region multiplied by the data rate consumed by each UE.

$$D_c = \sum_{z=1}^{N_z} \bar{U}_z D_{GBR} \quad (17)$$

## V. NUMERICAL RESULTS

Since the state-of-the-art models for computing the UE blocking probability (see section II) are not appropriate for GBR slices under our assumptions (i.e., reduced UE mobility and arbitrary distribution for the duration of the UE sessions), we cannot provide a fair comparison between these models and our model. For this reason, in this section we only focus on the validation of the proposed model. First, we present the experimental setup. Then, we analyze the aspects that impact on the execution time of our model. Finally, we evaluate the relative error for the UE blocking probability. Due to space limit, this section is focused on this parameter. Notwithstanding, we have also derived similar results for the mean number of consumed RBs and the cell capacity.

### A. Experimental Setup

To validate the proposed model, we use a Matlab-based simulator that simulates the arrival and departure of UE sessions in a cell. This simulator receives the PDF for the average SINR  $f_{PDF}(\bar{\gamma})$  as an input parameter. Table I summarizes the configuration parameters. Regarding the access technology, we assume a 5G New Radio (5G-NR) cell implementing an OFDMA scheme with  $\Delta f = 15$  KHZ, and  $N_{SC} = 12$ . We also consider several cell bandwidths from 10 MHz to 20

TABLE I  
 CONFIGURATION PARAMETERS

Parameters	Configuration
Access Technology	5G NR
Sub-carrier Spacing $\Delta f$ (OFDMA)	15 KHz
Sub-carriers per RB $N_{SC}$ (OFDMA)	12
Access Node: Bandwidth $W$ / Number of RBs $N_{RB}^{slot}$	10 MHz / 52 RBs; 15 MHz / 79 RBs; 20 MHz / 106 RBs;
PDF average SINR in the cell: $f_{PDF}(\bar{\gamma})$	Built using a dataset from a live LTE network
Service GBR $D_{GBR}$	10 Mbps (e.g., on demand HD videostreaming)
PDF Regions	5, 9, 15 linearly spaced regions
Traffic Load $\rho$	From 0.2 to 1.5

MHz [14]. Additionally, the radio resource allocation in 5G-NR is carried out by multiples of 2, 4, 8 and 16 RBs [15]. In our evaluation, we consider each UE consumes multiples of 4 RBs, thus Eq. 6 was accordingly modified. With respect to  $f_{PDF}(\bar{\gamma})$ , we have derived it by using real dataset from a Long Term Evolution (LTE) network. Note that 5G dataset is not available due to the deployment of 5G networks are already in an early stage. Specifically, this dataset contained the probabilities of reporting a certain Channel Quality Indicator (CQI). This means we have directly used the Table 5.2.2.1-3 in [15] to map each probability into the spectral efficiency achieved by a specific CQI, i.e., this table correspond to the spectral efficiency provided by the 5G New Radio (NR) standard. For the GBR service, we have assumed a data rate of  $D_{GBR} = 10$  Mbps.

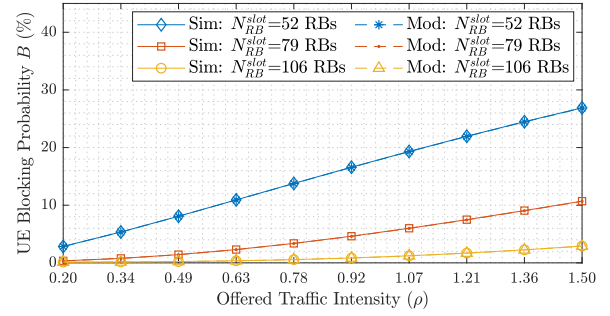
Considering these configuration parameters, we have evaluated the UE blocking probability in function of the offered traffic intensity. Specifically from  $\rho = 0.2$  to 1.5. Additionally, we have considered different values for the number of regions, from  $N_Z = 5$  to 15. All the experiments have been carried out on a computer with 16 GB RAM and an Intel core i7-7700HQ @ 2.80 GHz.

### B. Execution Time Evaluation

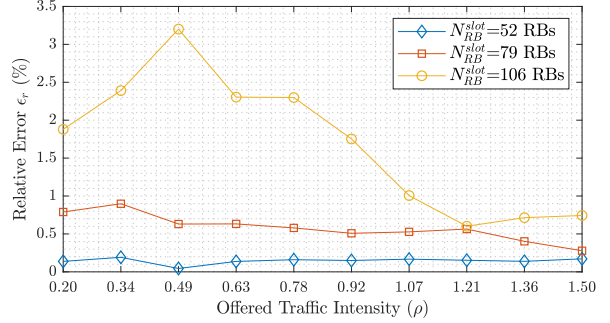
We have assessed the time complexity of our analytical model in two scenarios. In the former, we have covered several cell bandwidths, considering  $N_Z = 9$ . In the latter, we have considered different number of regions, with a cell bandwidth of 15 MHz (i.e., 79 RBs). The results for both scenarios are shown in Table II. We observe the execution time grows exponentially with the number of regions and cell bandwidth. The reason is using higher values for both parameters involves an increment in the number of states in the Markov chain as Eq. 8 shows. Note that the execution time does not depend on the offered traffic intensity.

 TABLE II  
 EXECUTION TIME

$N_Z = 9$			$N_{RB}^{slot} = 79$		
$N_{RB}^{slot} = 52$	$N_{RB}^{slot} = 79$	$N_{RB}^{slot} = 106$	$N_Z = 5$	$N_Z = 9$	$N_Z = 15$
0.009 s	0.0036 s	0.171 s	0.006 s	0.032 s	0.343 s



(a) UE Blocking Probability: Model vs Simulation



(b) Relative error

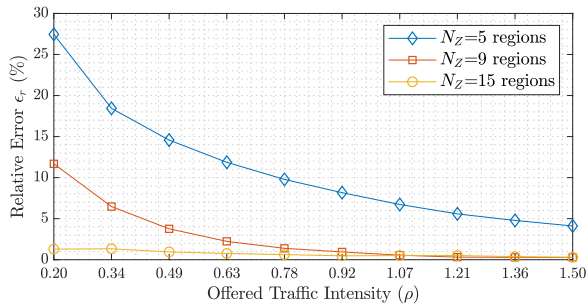
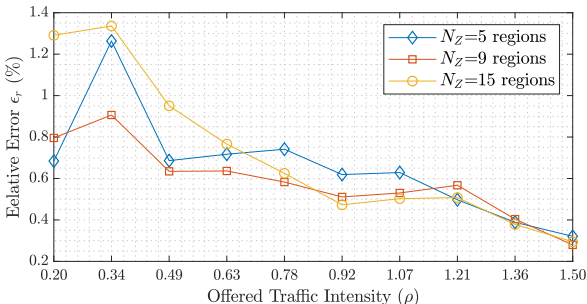
Fig. 4. Evaluation of UE Blocking Probability for different cell bandwidths.

### C. Model Validation

To validate our model, we have computed the relative error as  $\epsilon_r(\%) = \frac{B_{sim} - B_{mod}}{B_{sim}} \cdot 100$ , where  $B_{sim}$  and  $B_{mod}$  denote the UE blocking probability extracted from the model and simulator, respectively.

In Fig. 4(a), we depict the UE blocking probability derived from our model and the simulator. It shows how the UE blocking probability increases when (a) the available RBs in the cell are decreased and (b) the offered traffic intensity increases. This graph is useful for network operators because it allows to decide the bandwidth for each cell (i.e.,  $N_{RB}^{slot}$ ) while a threshold for  $B$  is provided, given certain conditions for the offered traffic intensity and the cell interference (i.e., a specific  $f_{PDF}(\bar{\gamma})$ ). Due to the scale used for the vertical and horizontal axes in Fig. 4(a), the error between the simulation and the model cannot be observed. In Fig. 4(b), we represent the relative error, which is below 3.5 % for any case. We also notice this error is higher with higher cell bandwidths. This fact is induced by the simulator because it takes less samples for the highest states (e.g., blocking states) when the number of states increases. The reason is the probability of reaching the blocking states is lower (i.e., see Fig. 4(a)), thus less data samples for these states are taken during the simulation (i.e., number of times in a blocking state).

Finally, we evaluate the relative error when our model use a different number of regions. In 5(a), each case were compared with the simulator implementing  $N_Z = 15$ . We observe  $\epsilon_r$  is below 1.5% when the model also implements  $N_Z = 15$ .


 (a) The simulator implements  $N_Z = 15$ 


(b) The simulator implements the same number of regions as the model

Fig. 5. Relative error in the evaluation of UE Blocking Probability for different number of regions

However, this error increases when the number of regions decreases. This increment is not induced by our model but rather the fact of splitting the  $f_{PDF}(\bar{\gamma})$  (see Fig. 1). When the number of regions is reduced and this number is small enough, each region could take a portion of the  $f_{PDF}(\bar{\gamma})$  different from the case of using a higher number of regions. This results in a slightly different distribution of  $\pi_z$ , which involves an unfair comparison. In Fig. 5(b), we observe the relative error considerably decreases when each simulation implements the same number of regions as our model, proving it suits the simulator results.

## VI. CONCLUSIONS AND FUTURE WORK

Network slicing is envisioned as a solution for providing emerging services over a common network infrastructure. Implemented as slices, most of these services will rely on data transmission with a strict GBR. Designing strategies for planning and operating GBR services could involve inherent issues such as the under(over)-provisioning of radio resources. To avoid that, it is crucial to model the UE blocking probability in each cell. Under this context, we propose an analytical model to evaluate this parameter. The main novelty is the consideration of a multi-dimensional Erlang-B system, which meets the reversibility property. This means our model is valid for arbitrary distributions of the UE session duration. This property also reduces the computation complexity of the model due to the solution for the state probabilities has product form. Additionally, our proposal considers the PDF for the average SINR in the cell to model the distribution of the channel

quality. The results show that our model exhibits an estimation error for the UE blocking probability below 3.5%.

Regarding the future work, several challenges lie ahead. One challenge is to include the effect of the scheduling discipline (e.g., proportional fair). Another challenge is considering services which simultaneously support VBR and GBR traffic.

## APPENDIX A

### REVERSIBILITY IN A MARKOV PROCESS

To proof the reversibility property of the proposed model, we follow the Kolmogorov cycle criteria [13]. This states that a necessary and sufficient condition for reversibility of a multi-dimensional Markov process is that for each dimension-pair, the circulation flow among four neighboring states in a square equals to zero (i.e., flow clockwise = flow counter-clockwise).

Considering four neighbor states from two arbitrary regions  $z$  and  $x$  (i.e.,  $z \neq x$ ):  $s_1 = (U_1, \dots, U_z, \dots, U_x, \dots, U_{N_z})$ ,  $s_2 = (U_1, \dots, U_z, \dots, U_x + 1, \dots, U_{N_z})$ ,  $s_3 = (U_1, \dots, U_z + 1, \dots, U_x + 1, \dots, U_{N_z})$ , and  $s_4 = (U_1, \dots, U_z + 1, \dots, U_x, \dots, U_{N_z})$ , we derive the clockwise and counter clockwise flows  $f_{cw} = \lambda_x \cdot p_1 \cdot \lambda_z \cdot p_2 \cdot (U_x + 1) \mu \cdot p_3 \cdot (U_z + 1) \mu \cdot p_4$  and  $f_{ccw} = \lambda_z \cdot p_1 \cdot \lambda_x \cdot p_4 \cdot (U_z + 1) \mu \cdot p_3 \cdot (U_x + 1) \mu \cdot p_4$ , respectively. We denote  $p_y$  the probability of state  $s_y$ . If we compare both equations, we easily check that the clockwise and the counter clockwise flows are equal. Thus, the proposed Erlang-B model is reversible.

## REFERENCES

- [1] P. Muñoz *et al.*, "Radio Access Network Slicing Strategies at Spectrum Planning Level in 5G and Beyond," *IEEE Access*, vol. 8, pp. 79604–79618, 2020.
- [2] M. P. Mota *et al.*, "Adaptive Modulation and Coding Based on Reinforcement Learning for 5G Networks," *IEEE Globecom*, pp. 1–6, 2019.
- [3] J. Navarro-Ortiz *et al.*, "A Survey on 5G Usage Scenarios and Traffic Models," *IEEE Commun. Surveys Tuts*, vol. 22, no. 2, pp. 905–929, 2020.
- [4] T. Bonald and A. Proutière, "Wireless Downlink Data Channels: User Performance and Cell Dimensioning," in *MobiCom*, pp. 339–352, 2003.
- [5] T. Bonald *et al.*, "Flow-level performance and capacity of wireless networks with user mobility," *Queueing Systems*, vol. 63, no. 1-4, p. 131, 2009.
- [6] S.-E. Elayoubi and T. Chahed, "Admission Control in the Downlink of WCDMA/UMTS," in *EuroNGI*, pp. 136–151, Springer, 2004.
- [7] D. K. Kim *et al.*, "A novel ring-based performance analysis for call admission control in wireless networks," *IEEE Commun. Lett.*, vol. 14, no. 4, pp. 324–326, 2010.
- [8] B. Sas *et al.*, "Modelling the time-varying cell capacity in LTE networks," *Telecommunication Systems*, vol. 55, no. 2, pp. 299–313, 2014.
- [9] C. Tarhini and T. Chahed, "QoS-oriented resource allocation for streaming flows in IEEE802.16e Mobile WiMAX," *Telecommunication Systems*, vol. 51, no. 1, pp. 65–71, 2012.
- [10] A. Abdollahpouri and B. E. Wolfinger, "Measures to quantify the gain of multicast with application to IPTV transmissions via WiMAX networks," *Telecommunication Systems*, vol. 55, no. 2, pp. 185–198, 2014.
- [11] M. Li, "Queueing Analysis of Unicast IPTV With Adaptive Modulation and Coding in Wireless Cellular Networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 10, pp. 9241–9253, 2017.
- [12] C. Kim *et al.*, "Mathematical Models for the Operation of a Cell With Bandwidth Sharing and Moving Users," *IEEE Trans. Wireless Commun.*, vol. 19, no. 2, pp. 744–755, 2020.
- [13] V. B. Iversen, "Teletraffic engineering and network planning," 2015.
- [14] 3GPP TS 38.101-1 V.16.3.0, "User Equipment (UE) radio transmission and reception; Part 1: Range 1 Standalone (Release 16)," Mar. 2020.
- [15] 3GPP TS 38.214 V.16.1.0, "NR, Physical layer procedures for data (Release 16)," Mar. 2020.