Multi-Connectivity in 5G New Radio: Optimal Resource Allocation for Split Bearer and Data Duplication

Jocelyne Elias[1], Fabio Martignon[2], and Stefano Paris[3]

[1]University of Bologna, jocelyne.elias@unibo.it
[2]University of Bergamo, fabio.martignon@unibg.it
[3]Nokia Bell Labs, stefano.paris@nokia-bell-labs.com

***Abstract*—Mobile radio networks have been evolving towards the integration of services and devices with a diverse set of throughput, latency, and reliability requirements. To support these requirements, 3GPP has introduced Multi Connectivity (MC) as a more flexible architecture for 5G New Radio (NR), where multiple radio links can be simultaneously activated to split or duplicate data traffic. Multi connectivity improves single user performance at the cost of higher interference due to the increase of radio transmissions, which negatively affects system throughput.**

This paper analyzes the problem of admission control and resource allocation in multi connectivity scenarios, considering different requirements and 5G NR features. Specifically, we formulate two optimization problems that leverage the features of the Packet Data Convergence Protocol (PDCP) layer, which controls the flow of data packets of the data radio bearer: the PDCP Split-Bearer Decision (PSD) and the PDCP Duplication Decision (PDD) problems, which are tailored for the enhanced Mobile Broadband (eMBB) and Ultra Reliable Low Latency Communications (uRLLC) services, respectively. We further provide heuristic approaches, specifically designed for the PSD and PDD problems, to effectively solve both these problems.

Numerical results in realistic network deployments confirm that our solutions can effectively allocate radio resources increasing admission rate and system throughput, while guaranteeing the required reliability level.

5G, Multi-Connectivity, Split-Bearer, Data Duplication, Optimization.

I. Introduction

The proliferation of interactive services and new form factor devices as well as the paradigm shift of sectors like healthcare, manufacturing, and transportation towards interconnected systems are calling for the design of more complex architectures and resource management schemes for next generation mobile networks. Indeed, mobile radio networks are expected to play an essential role for the interconnection of new devices and the seamless integration of new services [1].

To satisfy the diverse set of throughput, latency, and reliability requirements, 5G protocols and standards have been rapidly evolving to make use of higher frequencies and larger spectrum, flexible waveforms and access methods, more sophisticated schemes for scheduling radio resources, as well as more flexible architectures to connect and host the elements of the mobile radio network [2]. For example, 5G New Radio (NR) has introduced Multi-Connectivity (MC) [3] as a simple and effective way for improving latency, reliability, and throughput of cellular communications. The concept has been further developed in 5G Advanced to improve scheduling capabilities across multiple connections. In MC a User Equipment (UE) is connected to two gNBs (gNodeB, which indicates a 5G wireless base station), each handling up to two cells configured on different carrier frequencies. Therefore, up to four cells can be used for data transmission in 5G NR when MC is activated. The need for handling a massive number of connections associated with larger data rate requirements of future services will require the increase of the number of cells hosted by a gNB. Hence the number of simultaneous connections used to serve a UE will likely extend beyond four links in future generations of wireless cellular systems. In MC, the Master gNB (MgNB) establishes the main control and signaling connection with the UE, and it can activate a Secondary gNB (SgNB) to set up auxiliary data connections with the UE. The gNB with the best data connection to the UE between the MgNB and SgNB hosts the PDCP (Packet Data Convergence Protocol) entity that receives packets of the data radio bearer from the Core Network (CN). This node is called *hosting node* as opposed to the *assisting node*, which can be used to transmit either duplicated or a portion of the data traffic of the UE. Therefore, the PDCP entity of the hosting node controls the flow of packets of the data radio bearer by deciding which packets must be transmitted through the hosting and assisting nodes. The PDCP layer performs packet duplication at the transmitter, while it eliminates duplicated packets at the receiver using the PDCP sequence number [4].

Deciding which radio connections to activate to duplicate packets or how to split the flow of packets to serve a data radio bearer are key problems in the design of next-generation mobile networks. Indeed, the use of multiple connections and the ability to split the traffic across multiple cells enable the use of spare resources to reduce system blockage and improve reliability. In this paper we formulate and solve two distinct problems that arise with MC: (1) the PDCP Split-Bearer Decision (PSD) problem, which aims at increasing the capacity for eMBB services, and (2) the PDCP Duplication Decision

(PDD) problem, which aims at increasing the reliability of data transmission as required by the Ultra Reliable Low Latency Communications (uRLLC).

Specifically, the PSD problem consists in (i) deciding which users to serve and (ii) deciding whether and how to *split* the traffic of admitted users across multiple cells (also referred to as *legs*) to meet the bandwidth requirements of the services. This problem has been firstly investigated in [5] and extended in this work to include user reliability requirements. On the other hand, the PDD problem decides (i) which users to serve and (ii) the subset of legs over which traffic is *duplicated* to improve reliability. In both models, all key features of 5G NR are captured and modeled. We first propose effective and exact approaches using Mixed-Integer Linear Programming (MILP) formulations that find the optimal solution. We then propose two heuristics based on decomposition and greedy schemes for solving respectively the PSD and PDD problems, which, coupled with a carefully chosen bound to the set of legs that each user can exploit, permit to consistently reduce the computing time while still achieving close-to-optimum solutions, even in real-size network scenarios.

Numerical results obtained in realistic mobile network deployments and traffic scenarios as defined in 3GPP [6] show the effectiveness of the proposed models and approaches. Our analysis permits to capture and quantify the trade-off that mobile operators must face between admitting more users and providing them the necessary resources to fulfill their requirements.

This paper is structured as follows: Section II discusses related work, while Section III presents the system model we consider for the formulation of our problems. Sections III-A and III-B illustrate the proposed mathematical (MILP) formulations of the PSD and PDD problems, respectively. Section IV presents the decomposition and heuristic approaches we propose to diminish consistently the time necessary to compute a close-to-the-optimum solution. Numerical results obtained in realistic mobile network scenarios, with a large number of gNBs and mobile devices, are presented and discussed in Section V. Finally, conclusions are illustrated in Section VI.

A. Abbreviations and Acronyms

Table I lists the abbreviations used throughout this paper.

Abbreviation	Description
5G	Fifth-generation of mobile technologies
BLER	BLock Error Rate
CN	Core Network
CQI	Channel Quality Indicator
CSI	Channel State Information
eMBB	enhanced Mobile Broadband
gNBs	gNodeB (5G wireless base station)
MC	Multi Connectivity
MCS	Modulation and Coding Scheme
MgNB	Master gNodeB
NR	New Radio
PDCP	Packet Data Convergence Protocol
PDD	PDCP Duplication Decision problem
PRB	Physical Resource Block
PSD	PDCP Split-Bearer Decision problem
RAN	Radio Access Network
SDAP	Service Data Adaptation Protocol
SgNB	Secondary gNodeB
SINR	Signal to Interference plus Noise Ratio
TTI	Transmission Time Interval
UE	User Equipment
uRLLC	ultra Reliable Low Latency Communications

TABLE I: Main abbreviations and acronyms.

II. Related Work

Multi-Connectivity has been proposed as a simple and effective way to satisfy data rates, latency, reliability and coverage requirements of 5G networks. Different options for connecting to multiple radio cells, including PDCP solutions, are discussed in [7], [8]. The possibility to enable simultaneous connections to multiple and distinct access points reduces drastically the outage probability especially during handover [9] for highly mobile users. In [10], authors show that a small number of connections (i.e., up to 4) improves both outage and spectral efficiency metrics in moderately dense network deployments. To simplify the evaluation of multi-connectivity techniques beyond outage probability and include throughput, latency and reliability metrics, authors in [11] present a closed-form expression to derive the symbol error rate as a function of the received SINRs across multiple connections.

Machine learning techniques have been proposed to reduce the complexity of scheduling radio resource in ultra-dense scenarios and improve network performance when multi connectivity is enabled [12], [13]. Furthermore, early attempts to enhance positioning schemes in mmWave deployments using MC have been discussed in [14], [15]. The work in [14] analyzes the power allocation problem across multiple connections to improve positioning accuracy, while [15] presents a scheme that improves the user positioning and base station orientation uncertainty by exploiting reference signals from multiple connections.

5G NR has specified PDCP duplication on top of multi connectivity as a simple method to improve latency and reliability of data transmission by exploiting the spatial, temporal and frequency diversity offered by multiple cells configured on different carrier frequencies [4], [16]. An analytical evaluation of the outage probability and resource utilization of multi-connectivity with PDCP data duplication is presented in [17]. Authors in [18] formulate PDCP data duplication as a mathematical optimization problem with latency and reliability constraints, and propose a heuristic approach to solve the problem. Another heuristic scheme is used in [19] to dynamically select data duplication only for users whose latency requirements are critical. In [20] authors evaluate the problem of selecting a subset of connections as well as the modulation and coding schemes (MCS) and the appropriate decoding error target for each enabled connection. The problem is formulated as a mixed integer non-linear program and solved with a combination of Newton and branch-and-bound methods.

Differently from data duplication, 5G NR has introduced split-bearer as a scheme to increase the throughput of data communications by splitting user traffic across multiple con-

nections. A control scheme to dynamically select the best subset of connections for data transmission using channel state information and cell load is presented in [21], while a method based on utility maximization to schedule resources among multiple user devices is proposed in [22]. Both works assume that all UEs can be admitted in the system and their traffic fully satisfied. Therefore, they do not consider the admission problem. Two enhancements involving user signaling to selectively duplicate transmissions of only lost packets have been evaluated in [23], [24]. Using any of these two schemes, a 5G system can serve higher loads without affecting the reliability of user transmissions. In [25] authors present a scheme to dynamically adjust the split ratios of users traffic across multiple connections in order to fulfill QoS requirements. Similarly, in [26], [27] authors present online control policies based on Lyapunov optimization to decide the split ratios and cell group state for downlink transmissions, while in [28] they provide a control scheme to select active links and allocate power for uplink transmissions.

A system-level analysis of split bearer and data duplication on top of multi-connectivity in 5G systems has been presented in [29], [30]. Both works show that the use of split bearer and data duplication permit to fulfill the throughput and reliability requirements of eMBB and URLLC services, respectively.

Differently from prior art, we formulate the joint admission control and resource allocation problem for multi-connectivity scenarios (both Split-Bearer and Duplication Decision problems) with exact mathematical formulations, and further propose heuristic approaches to speed up the computation of the solution.

III. System Model and Problem Formulations for PSD and PDD

In this section, we describe the system model that we consider in this paper. Then, we present the mathematical formulation of two optimization problems: the PDCP Split-Bearer Decision (PSD, Section III-A) and the PDCP Duplication Decision (PDD, Section III-B) problems.

We model the Radio Access Network (RAN) as a bipartite graph $G = (\mathcal{U}, \mathcal{L}, \mathcal{L}_u)$, where $\mathcal{U}$ and $\mathcal{L}$ represent the set of users (UEs) and the set of all cells, respectively. $\mathcal{L}_u \subseteq \mathcal{L}$ models the set of cells that can be used by UE $u \in \mathcal{U}$ to transmit a data packet ($\mathcal{L} = \bigcup_{u \in \mathcal{U}} \mathcal{L}_u$), while $\mathcal{L}_l \subseteq \mathcal{L}$ defines the set of all legs that interfere with leg $l \in \mathcal{L} \setminus \{l\}$.

Since the connection between a UE and a cell uniquely identifies a transmission leg, each element $l \in \mathcal{L}_u$ is called indifferently *leg* or *cell* in the rest of the paper. Among all legs in $\mathcal{L}_u$ that can be used to transmit a data packet, one leg is used to control the UE u. We identify this primary leg using the function $l_0(u) : \mathcal{U} \to \mathcal{L}$.

The system bandwidth, numerology, number of symbols per slot and number of subcarriers per physical resource Block (PRB) are denoted as B, μ, N_s, and N_c, respectively. The numerology defines the subcarrier spacing (SCS), B_μ, and the number of slots in a second, N_μ. G_{ul} models the channel gain between a cell $l \in \mathcal{L}$ and user $u \in \mathcal{U}$, while P_{ul} represents the power used by a cell to serve the UE. We observe that the set of cells that can be used by a UE, $\mathcal{L}_u$, is estimated based on the user placement and received signal power, $G_{ul}P_{ul}$. In order to speed up the computation of a solution, some of the legs can be pruned by simply applying a threshold on the received signal power strength. During the network operation, user measurements can be also collected and the set can be further refined. For a downlink transmission, the cell selects the highest Modulation and Coding Scheme (MCS) to match the target error probability according to the Signal to Interference plus Noise Ratio (SINR). We model the list of available MCSs as the discrete set $\mathcal{M}$. For each MCS $m \in \mathcal{M}$, R_m, Q_m and ρ_m identify the transmission rate, the modulation order and the code rate, respectively. Since in 5G systems the smallest scheduling granularity is the Physical Resource Block (PRB), all system parameters are scaled accordingly. We represent the range of SINRs at which UEs operate as a discrete set whose indices are defined in the set $\mathcal{S}$. For each SINR index $s \in \mathcal{S}$, the real value $\gamma_s \in \mathbb{R}$ represents the corresponding SINR value. In other words, $\gamma_s : \mathcal{S} \to \mathbb{R}$ defines the function that maps SINR indices to SINR values. The discretization step of the SINR range is an input parameter of our models; it can be defined so that the elements of $\mathcal{S}$ match the possible values of the Channel Quality Indicator (CQI) periodically reported by the UE during the monitoring of the channel. Finally, the BLock Error Rate (BLER), which corresponds to the error probability a data packet experiences when it is transmitted using MCS $m \in \mathcal{M}$ at SINR $s \in \mathcal{S}$, is indicated by the real value ϕ_{ms}.

The RAN must serve the traffic of each user $u \in \mathcal{U}$ fulfilling the data rate D_u of the data connection with a certain reliability target Φ_u. Assuming a continuous stream for the traffic flow, the data rate D_u expressed in bit per seconds can be re-scaled into *bit per slot*, which corresponds to the decision timescale of the scheduler: $\frac{D_u}{N_\mu}$. We underline that the algorithms we present to solve the PSD and PDD problems operate in the non-real time RAN Intelligent Controller (RIC), which is the only network element with the complete network view needed for activating legs to users. Physical resources are allocated by the scheduler operating in the real-time controller of the gNB. Nevertheless, capacity constraint must be taken into consideration by the non-real time RIC when user legs are activated.

The energy consumption of a user device configured in multi-connectivity increases proportionally to the number of active legs used to connect to multiple cells. The larger energy consumption is caused by the activation of additional dedicated antennas, Radio-Frequency (RF) chains and Baseband (BB) processing for each cell. Additionally, the user device needs to monitor a larger bandwidth and report channel state information (CSI) for all active legs. High power consumption puts additional strain on the user device since it reduces battery lifetime. Therefore, the optimization of the network configuration needs to include the minimization of the consumed power in addition to the user service rate. To this end, we model the power P_u consumed by a user device u when configured with multiple active legs according to the 3GPP model [31]

Parameter	Description
$\mathcal{U}$	Set of users (UEs)
$\mathcal{L}$	Set of legs
$\mathcal{M}$	Set of Modulation and Coding Schemes
$\mathcal{S}$	Set of indices of the discrete SINR values
$\mathcal{U}_l$	UEs that can be served by leg $l \in \mathcal{L}$
$\mathcal{L}_u$	Set of legs that can be used to serve UE u
$\mathcal{L}_l$	Legs that interfere with leg $l \in \mathcal{L}$
$l_0(u) : \mathcal{U} \to \mathcal{L}$	Function to identify UE's primary leg
P_{ul}	Power used by leg $l \in \mathcal{L}$ to serve UE $u \in \mathcal{U}$
G_{ul}	Channel gain for UE $u \in \mathcal{U}$ on leg $l \in \mathcal{L}$
$\mathcal{N}_0$	Noise
γ_s	SINR value corresponding to index $s \in \mathcal{S}$
ϕ_{ms}	BLER (BLock Error Rate) when using MCS $m \in \mathcal{M}$ with SINR γ_s ($s \in \mathcal{S}$)
R_m	Rate for MCS $m \in \mathcal{M}$ [bit/s]
Q_m	Number of modulated bits of MCS $m \in \mathcal{M}$ [bit/symbol]
ρ_m	Code rate of MCS $m \in \mathcal{M}$
D_u	Traffic data rate of UE u [bit/s]
Φ_u	BLER target for UE u
α	weight parameter between # of accepted users and # of used legs
λ_u	Fraction of time UE u spends in active mode
L_u	Number of active legs used to serve UE u
B	System bandwidth
N_c	Number of subcarriers per PRB
N_s	Number of symbols per slot
B_μ	Subcarrier spacing (SCS)
N_μ	Number of slots per second
P_u^r	Consumed power in active mode
P_u^s	Consumed power in sleep mode

Variable	Description
$x_{ulms} \in \{0,1\}$	Indicates whether MCS m of leg l is used to serve UE u when the SINR index is s
$y_{ul} \in \{0,1\}$	Auxiliary variable indicating whether leg l is used to serve UE u
$z_u \in \{0,1\}$	Indicates whether UE u is admitted

TABLE II: Parameters and decision variables used in our system model and mathematical formulation of the PSD and PDD problems.

as follows:

$$P_u = \begin{cases} P_u^r \lambda_u + P_u^s (1 - \lambda_u) & L_u = 1 \\ 0.85 L_u \left(P_u^r \lambda_u + P_u^s (1 - \lambda_u)\right) & L_u > 1, \end{cases} \quad (1)$$

where the term λ_u represents the fraction of time a user device spends in active mode, while the term L_u is the number of active legs used to serve the user device. Terms P_u^r and P_u^s represent the power consumed by the user in active and sleep modes, respectively, as defined by 3GPP [31]. We observe that the term L_u can be computed from the problem parameters and decision variables as described below. In contrast, the term λ_u depends on the scheduling policy and the traffic flow. In our case, $\lambda_u = 1$ since we assume continuous traffic stream and all users eligible for scheduling in each slot.

A simplified RAN deployment with the main parameters of our system model is illustrated in Figure 1. The RAN is composed of two UEs and three cells (one per gNB). The cell l_1 handled by gNB_1 is the primary leg for UE u, whereas cell l_3 of gNB_3 is the primary leg for UE i. Both UEs share the cell handled by gNB_2 as secondary leg. Since cells of gNB_1 and gNB_3 are configured on the same carrier frequency, they may interfere with each other. Therefore, transmissions on legs $(u; l_1)$ and $(i; l_3)$ must be coordinated to avoid cross interference. We can observe that when split bearer is configured as illustrated in Figure 1(a), the primary and secondary legs are used for the transmission of two different packets (A and B for UE_u), while in data duplication the same packet (A) is transmitted across all legs as depicted in Figure 1(b). Therefore, split bearer increases the available *throughput*, whereas data duplication improves transmission *reliability*.

Binary variables x_{ulms} indicate whether leg l is used to serve user u with MCS m when the reported SINR index is s (i.e., $x_{ulms} = 1$). In both formulations the first leg represents the primary leg, namely the best leg connecting the UE u to its MgNB. This leg is used to exchange control information between the MgNB and the UE in addition to data traffic. Binary variables y_{ul} are auxiliary variables indicate whether user leg l is used to serve user u. They permit to simplify the formulation of SINR constraints and the description of the optimization problem, but can be omitted when solving the problem to reduce the memory needed to store decision variables. Note that the number of active legs used to in the power model described above can be computed as $L_u = \sum_{l \in \mathcal{L}_u} y_{ul}$. Finally, binary variables z_u indicate whether the traffic of UE u is fully served by the network (i.e., $z_u = 1$). All parameters and decision variables used in our mathematical formulations are summarized in Table II. The parameters represent the state of the network at decision time.

A. PDCP Split Bearer Decision Problem

We now formulate the PSD problem, which consists in deciding whether and how to split the traffic across multiple legs to meet the bandwidth requirements of user services. This model is specifically tailored for implementing the eMBB (enhanced Mobile Broadband) service in a mobile network. The PSD problem can be formulated as the optimization problem (2)-(12):

$$\max \sum_{u \in \mathcal{U}} z_u - \alpha \sum_{u \in \mathcal{U}, l \in \mathcal{L}_u} y_{ul} \qquad \text{s. t.:} \quad (2)$$

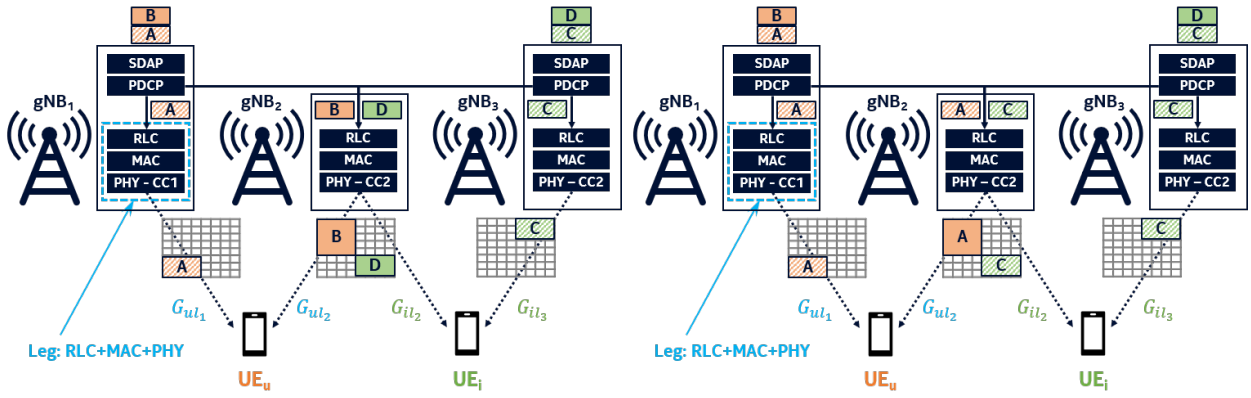
$$\sum_{l \in \mathcal{L}_u} \sum_{m \in \mathcal{M}} \sum_{s \in \mathcal{S}} x_{ulms} \geq z_u \qquad \forall u \in \mathcal{U}, \quad (3)$$

$$\sum_{l \in \mathcal{L}_u} y_{ul} \geq z_u \qquad \forall u \in \mathcal{U}, \quad (4)$$

$$\sum_{m \in \mathcal{M}} \sum_{s \in \mathcal{S}} x_{ul_0(u)ms} \leq y_{ul_0(u)} \qquad \forall u \in \mathcal{U}, \quad (5)$$

$$\sum_{m \in \mathcal{M}} \sum_{s \in \mathcal{S}} x_{ulms} \leq y_{ul} \qquad \forall u \in \mathcal{U}, l \in \mathcal{L}_u \setminus \{l_0(u)\}, \quad (6)$$

$$y_{ul_0(u)} |\mathcal{L}| \geq \sum_{l \in \mathcal{L}_u \setminus \{l_0(u)\}} y_{ul} \qquad \forall u \in \mathcal{U}, \quad (7)$$



(a) PDCP Split Bearer

(b) PDCP Data Duplication

Fig. 1: Example of RAN with two UEs and three cells. Each UE activates two legs. The gNB_2 node is the SgNB and assisting node of both gNB_1 and gNB_3, whihc are MgNBs and hosting nodes of for UE_u and UE_i, respectively. When the Split Bearer is configured (a): the primary and secondary legs are used for the transmission of two different packets (A and B for UE_u). In Data Duplication (b): the same packet (A for UE_u) is transmitted across all legs.

$$\frac{P_{ul}G_{ul}}{\mathcal{N}_0 + \sum_{j\in\mathcal{L}_l, k\in\mathcal{U}_j : k\neq u} P_{kj}G_{uj}y_{kj}} \geq \gamma_s x_{ulms} \qquad \forall u \in \mathcal{U}, l \in \mathcal{L}_u, m \in \mathcal{M}, s \in \mathcal{S}, \tag{8}$$

$$\sum_{l\in\mathcal{L}_u}\sum_{m\in\mathcal{M}}\sum_{s\in\mathcal{S}} R_m x_{ulm} \geq D_u z_u \qquad \forall u \in \mathcal{U}, \tag{9}$$

$$-\sum_{m\in\mathcal{M}}\sum_{s\in\mathcal{S}} log_{10}\left(\phi_{ms}\right) x_{ulms} \geq -log_{10}\left(\Phi_u\right) z_u, \qquad \forall u \in \mathcal{U}, l \in \mathcal{L}_u, \tag{10}$$

$$\sum_{u\in\mathcal{U}_l}\sum_{m\in\mathcal{M}}\sum_{s\in\mathcal{S}} \left\lceil \frac{R_m}{N_\mu N_s N_c Q_m \rho_m} \right\rceil x_{ulms} \leq \frac{B}{N_c B_\mu} \qquad \forall l \in \mathcal{L}, \tag{11}$$

$$x_{ulms}, y_{ul}, z_u \in \{0,1\} \qquad \forall u \in \mathcal{U}, l \in \mathcal{L}_u, m \in \mathcal{M}, s \in \mathcal{S}. \tag{12}$$

The optimization function (2) maximizes the number of served UEs, and pursues also the minimization of the used resources, α being the weight that permits to trade-off between the two objectives. Constraints (3) and (4) state that at least one leg must be activated with a given MCS to serve user u when such user is admitted (i.e., when $z_u = 1$). Constraints (5) and (6) force the use of at most one MCS for the primary leg and the secondary leg, respectively. Note that if UE u is not served, no MCS is selected and all corresponding variables are set to zero. Constraints (7) force the activation of the primary leg if any secondary leg is used to transmit part of the traffic of UE u. If the primary leg is not used, the traffic cannot be split on any other secondary leg. Constraints (8) model the Signal to Interference plus Noise Ratio (SINR) perceived by any UE u for leg l. A constraint is defined for each MCS according to its SINR index. We observe that any MCS that satisfies the SINR threshold can be selected to serve a user u. However, constraints (5) and (6) limit the choice to a single value. Constraints (9) guarantee that the aggregated rate of all activated legs is enough to satisfy the traffic demand D_u of user u, while constraints (10) guarantee that each activated leg uses a MCS that can support the target BLER Φ_u defined for user u. The set of constraints (11) are the capacity constraints. The right-hand-side of the inequality represents the number of PRBs available in a slot, whereas the left-hand-side accounts for the number of PRBs the scheduler must allocate to serve all users connected to leg l. The ratio $\frac{R_m}{N_\mu}$ indicates the portion of user bitrate per slot served by leg l, while the product $N_s N_c Q_m \rho_m$ corresponds to the bits that can be transmitted using MCS m. Note that R_m is an upper bound of the user data rate D_u and the user may be overprovisioned. Finally, the set of constraints (12) defines the range of the decision variables.

B. PDCP Duplication Decision Problem

The *PDCP Duplication Decision (PDD)* problem consists in deciding whether to duplicate traffic over multiple legs and which subset of legs to use for the data transmission. Duplicating data packets increases transmission opportunities over different channels that exhibit different temporal and spatial fading conditions. This increases the reliability of data transmission against adverse channel conditions that is particularly important for Ultra Reliable Low Latency Communications (URLLC). More specifically, assuming independent packet losses across legs, the loss probability of a packet duplicated on L legs is equal to $\prod_{i=1}^{L} p_i$, where p_i is the loss probability of leg l. Assuming the packet is transmitted using MCS m on a leg that experiences SINR γ_s, the leg loss probability is $p_i = \phi_{ms}$. Therefore, the packet loss probability decreases steeply with the number of legs used for its transmission. However, the duplication of traffic across multiple legs results in higher intra-cell and inter-cell interference that in turn may end up negatively affecting the overall system performance. Therefore, a trade-off between user reliability and system throughput emerges for duplication decisions.

The PDD problem can be formulated as the optimization

problem (13)-(23):

$$\max \quad \sum_{u\in\mathcal{U}} z_u - \alpha \sum_{u\in\mathcal{U}, l\in\mathcal{L}_u} y_{ul} \tag{13}$$

$$\text{s.t.:} \quad \sum_{l\in\mathcal{L}_u}\sum_{m\in\mathcal{M}}\sum_{s\in\mathcal{S}} x_{ulms} \geq z_u \qquad \forall u\in\mathcal{U}, \tag{14}$$

$$\sum_{l\in\mathcal{L}_u} y_{ul} \geq z_u \qquad \forall u\in\mathcal{U}, \tag{15}$$

$$\sum_{m\in\mathcal{M}}\sum_{s\in\mathcal{S}} x_{ul_0(u)ms} \leq y_{ul_0(u)} \qquad \forall u\in\mathcal{U}, \tag{16}$$

$$\sum_{m\in\mathcal{M}}\sum_{s\in\mathcal{S}} x_{ulms} \leq y_{ul} \qquad \forall u\in\mathcal{U}, l\in\mathcal{L}_u\setminus\{l_0(u)\}, \tag{17}$$

$$y_{ul_0(u)} \geq \sum_{l\in\mathcal{L}_u\setminus\{l_0(u)\}} y_{ul} \qquad \forall u\in\mathcal{U}, \tag{18}$$

$$\frac{P_{ul}G_{ul}}{\mathcal{N}_0 + \sum_{j\in\mathcal{L}_l}\sum_{k\in\mathcal{U}_j} P_{kj}G_{uj}y_{kj}} \geq \gamma_s x_{ulms} \quad \forall u\in\mathcal{U}, l\in\mathcal{L}_u, m\in\mathcal{M}, s\in\mathcal{S}, \tag{19}$$

$$\sum_{m\in\mathcal{M}}\sum_{s\in\mathcal{S}} R_m x_{ulms} \geq D_u y_{ul} \qquad \forall u\in\mathcal{U}, l\in\mathcal{L}_u, \tag{20}$$

$$-\sum_{l\in\mathcal{L}_u}\sum_{m\in\mathcal{M}}\sum_{s\in\mathcal{S}} log_{10}\left(\phi_{ms}\right) x_{ulms} \geq -log_{10}\left(\Phi_u\right) z_u \quad \forall u\in\mathcal{U}, \tag{21}$$

$$\sum_{u\in\mathcal{U}_l}\sum_{m\in\mathcal{M}}\sum_{s\in\mathcal{S}} \left\lceil \frac{D_u}{N_\mu N_s N_c Q_m \rho_m} \right\rceil x_{ulms} \leq \frac{B}{N_c B_\mu} \quad \forall l\in\mathcal{L}, \tag{22}$$

$$x_{ulms}, y_{ul}, z_u \in \{0,1\} \ \forall u\in\mathcal{U}, l\in\mathcal{L}_u, m\in\mathcal{M}, s\in\mathcal{S}. \tag{23}$$

The optimization function (13) and the set of constraints (14)-(19) correspond to the objective function (2) and constraints (3)-(8) of the PSD problem.

In contrast, the set of constraints (20) guarantees that each leg selected for user u can serve its traffic demand. The left-hand-side of the constraint defines the transmission rate used to serve user u on leg l, while the right-hand-side represents the requested user's data rate D_u. Note that the right-hand-side of the inequality is set to zero if the user is not admitted and the left-hand-side can take any value. Similarly, the set of constraints (21) guarantees that the target error probability of user u is met. Here, the left-hand-side of the constraint defines the joint error probability of all legs assuming independent errors, while the right-hand-side represents the BLER target for UE u. Constraints (20) and (21) represent therefore the QoS requirements of the users. The set of constraints (22) are the capacity constraints. The right-hand-side of the inequality represents the number of PRBs available in a slot, whereas the left-hand-side accounts for the number of PRBs the scheduler must allocate to serve all users connected to leg l. The ratio $\frac{D_u}{N_\mu}$ indicates the user bitrate per slot, while the product $N_s N_c Q_m \rho_m$ corresponds to the bits that can be transmitted using MCS m. Finally, constraints (23) define the range of the decision variables.

C. Refinements of PSD and PDD Models

We observe that the PSD (2)-(12) and PDD (13)-(23) models contain non-linear constraints and a trade-off constant α. Hereafter, we provide further details to refine the two models.

1) Linearization of SINR constraints: Constraints (8) and (19) in the PSD and PDD models are not linear. However, they can both be linearized by appropriately defining a large constant M as follows:

$$P_{ul}G_{ul}x_{ulms} + \left(1 - x_{ulms}\right) M \geq \left(\mathcal{N}_0 + \sum_{j\in\mathcal{L}_l}\sum_{k\in\mathcal{U}_j} P_{kj}G_{uj}y_{kj}\right)\gamma_s \ .$$

The constraints simply state that when a leg has been activated ($x_{ulms} = 1$), the received power at the UE must be larger than the sum of interference plus noise. In contrast, when the leg is not activated for any MCS and SINR index ($x_{ulms} = 0$), the big M value makes the constraint valid for any value of the interference generated by other legs.

2) Objective function and weighting parameter: The weighting constant α in objective functions (2) and (13) acts as a trade-off parameter between admission and resource utilization. If α is set equal to $\alpha = \frac{1}{|\mathcal{L}||\mathcal{U}|+1}$, the parameter imposes a lexicographic order between the admission and resource utilization objectives. In particular, we first optimize the number of accepted users and then we carefully allocate radio resources to satisfy their required needs. This is a typical radio planning choice that is aligned with the interests of mobile operators.

3) Latency requirements: In our problem formulations we explicitly model only the reliability constraints since reliability requirements of industrial services can be fulfilled only using multiple connections that increase interference and reduce available resources for other services. Latency requirements can be implicitly considered using the equivalent capacity concept introduced in [32], or explicitly modeled introducing additional constraints as follows:

$$\sum_{m\in\mathcal{M}}\sum_{s\in\mathcal{S}} \frac{S_u}{R_m} x_{ulm} \leq L_u z_u \qquad \forall u\in\mathcal{U}, l\in\mathcal{L}_u,$$

where S_u represents the size of a typical packet generated by the service of user u, while L_u is the maximum latency for the transmission of a packet for user u. These constraints force the use of MCSs that fulfill the latency requirement of the packet of size S_u for each leg used to serve user u. If the constraint cannot be satisfied for all selected legs, the user is not admitted.

IV. Algorithms for Solving the PSD and PDD Problems

The PSD and PDD problems described in the previous section are NP-hard since they both extend the classical generalized assignment problem. Therefore, the computing time necessary to solve them increases steeply with the instance size. Even medium-size instances as those illustrated in Section V require hours to be solved using state of the art

solvers. This is essentially due to the constants α and M that make relaxation techniques inefficient.

In order to reduce the overall complexity, we first propose in Section IV-A a *decomposition approach* tailored to the PSD problem, which consists in solving sequentially the *users' admission problem* and the *radio resource allocation problem*. Then, in Section IV-B we illustrate an efficient *greedy* algorithm to solve the PDD problem.

We observe that with the setting of the weighting constant α discussed in Section III.C.2, naturally adopted in the literature as well as in our paper, we can compute the optimal solution of the PSD problem within few hundreds of seconds, in the worst case, as illustrated in Section V (see for example Fig. 7), and in many cases in a much lower amount of time, so that further devising approximated or suboptimal/greedy approaches is less interesting in such case. On the other hand, the PDD problem is more complex since traffic is entirely duplicated over multiple legs (this second problem is more demanding in terms of resource allocations with respect to the first). The computing time hence increases in PDD, and a decomposition approach is not sufficient to compute solutions in a reasonable time. For these reasons we proposed a heuristic approach to obtain good solutions in a reasonable time.

A. Decomposition Approach for the PSD Problem

Algorithm 1 illustrates the main steps of the decomposition approach we propose to solve the PSD problem. In addition to the parameters described in Table II, the algorithm takes as input the maximum number of legs (L) each user can be connected to. The proposed approach firstly reduces the number of available legs that can be used to serve each user. Specifically, for each user u it keeps only the two legs with the strongest signal $P_{ul}G_{ul}$: $\mathcal{L'}_u = \{l_1, l_2\} \subseteq \mathcal{L}_u$, where $\forall l'' \in \mathcal{L}_u, l'' \neq l_1, l_2$ we have $P_{ul_1}G_{ul_1} > P_{ul''}G_{ul''}$, $P_{ul_2}G_{ul_2} > P_{ul''}G_{ul''}$, and $l_1 \neq l_2$. It then solves the problem (2)-(12) with the objective of maximizing the number of accepted users, $\sum_{u\in\mathcal{U}} z_u$. This step corresponds to solving the *admission problem* without optimizing the used radio resources and produces as output a vector of binary variables $\boldsymbol{z'}$.

Once the admission has been solved, an instance of the problem (2)-(12) is created with variables $\boldsymbol{z}$ already fixed according to the solution found in the admission step, namely $\boldsymbol{z} = \boldsymbol{z'}$. This new instance is then solved considering only the objective of minimizing the number of used legs $\sum_{u\in\mathcal{U}, l\in\mathcal{L}_u} y_{ul}$ (note that now the term $\sum_{u\in\mathcal{U}} z'_u$ in the objective function is a constant and can be omitted).

B. Greedy Algorithm for the PDD Problem

Algorithm 2 illustrates the main steps of the greedy algorithm we propose to solve the PDD problem. In addition to the system parameters, the algorithm gets two additional inputs: the maximum number of legs that can be used to serve each user (L), and the percentage of users that can be served with more than one leg η (i.e., the users that can be configured in multi-connectivity). This latter parameter permits to evaluate the effect of multi-connectivity on the amount of used resources and the system capacity in terms of number of accepted users. The initialization step 1 consists in sorting users according to a certain metric. This ranking simplifies the admission decision performed in the next steps. Specifically, we first compute the SINR γ_{ul} perceived by each user u on every leg l assuming that all users are fully connected using all their available legs. Then, we compute the combined SINR Γ_u for each user as the product of all SINR values across all legs that can be used to serve a user u (i.e., all legs in $\mathcal{L}_u$). Once the combined SINRs have been computed, we sort users in ascending order of the ratio between the data rate D_u and the combined SINR Γ_u. In this way, we first serve users with small data rate and good channel conditions. The rationale behind such choice is that these users are usually close to gNBs and generate low interference, hence their allocation is rather simple and slightly affects allocation decisions of users at the edge.

Algorithm 1: PSD Decomposition Scheme

Data: L: maximum number of legs
Result: Value of variables $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$
1 $\forall u \in \mathcal{U}$ keep in $\mathcal{L}_u$ the L legs with the highest $P_{ul}G_{ul}$ and remove all other legs;
$\mathcal{L} \leftarrow \bigcup_{u\in\mathcal{U}} \mathcal{L}_u$;
2 $\alpha \leftarrow 0$;
$\boldsymbol{z'} \leftarrow$ Solve problem (2)-(12);
3 Restore original value of α;
$(\boldsymbol{x}, \boldsymbol{y}) \leftarrow$ Solve problem (2)-(12) with $\boldsymbol{z} = \boldsymbol{z'}$;

Once users have been sorted, in step 2 for each user u we select a combination of legs that meets the user requirements in terms of data rate D_u and BLER target Φ_u. We also check whether the capacity of each leg is not exceeded. To this end, we first generate a set $\mathcal{T}_u$ that contains all possible combinations of legs of size smaller than or equal to the input parameter L. For each combination of legs $\mathcal{M}_u \in \mathcal{T}_u$, we first check if it can satisfy the user requirements and capacity constraints (step 3) as well as if it improves the solution (relation $\succ$ in step 4). If all conditions hold, the combination $\mathcal{M}_u$ is selected as new best candidate solution for user u if either the percentage of users that have activated more than one leg in the solution computed so far is smaller than η or it contains one leg.

To verify whether selecting a combination of legs $\mathcal{M}_u$ makes the final solution feasible we check both user requirements and capacity constraints. Specifically, we split the original BLER target Φ_u across all legs of the combination proportionally to their SINR. The BLER target of each leg l of the analyzed combination $\mathcal{M}_u$ is $\phi_{ul} = \frac{\Phi_u}{\gamma_{ul}}$, and we select for leg l the highest MCS $m \in \mathcal{M}$ that can tolerate the BLER ϕ_{ul}. With the selected MCSs, we further verify whether capacity constraints (22) hold trues for each $l \in \mathcal{M}_u$. If the corresponding rate R_m is larger than or equal to the user data rate D_u and the number of available PRBs is not exceeded, the variables of leg l are updated accordingly. Otherwise, the entire combination $\mathcal{M}_u$ is discarded, since all legs must serve the

user data rate D_u. Note that the combination $\mathcal{M}_u$ is selected in the final solution only if it is strictly better than another combination $\mathcal{B}_u$, which has been previously selected as best candidate solution for user u. In our implementation, we define the best combination as the one with the lowest number of legs, in order to limit resource utilization. Therefore, the relation $\mathcal{M}_u \succ \mathcal{B}_u$ is defined as $\mathcal{M}_u \succ \mathcal{B}_u \iff |\mathcal{M}_u| < |\mathcal{B}_u|$.

Finally, in step 5 we recompute the SINR of all legs according to the new solution and we reorder the remaining users. We observe that the deactivation of either some legs (i.e., $z_u = 1$ and $\sum_{l \in \mathcal{L}_u} y_{ul} \leq |\mathcal{L}_u|$) or all legs (i.e., $z_u = 0$) changes the interference perceived by users that have not been yet analyzed by the algorithm. Therefore, the SINR update and user reordering helps the next user allocation decisions.

We observe that, for each user u, Algorithm 2 scans a number of elements $n = \sum_{k=1}^{L} \binom{m}{k} k$, where $m = |\mathcal{L}_u|$, since the number of combinations of legs for each user u is $|\mathcal{T}_u| = \sum_{k=1}^{L} \binom{m}{k}$ and each combination $\mathcal{M}_u$ contains $k = 1, 2, ..., L$ legs. Therefore, the number of elements for each user u grows as $\frac{m^L}{(L-1)!}$. While the growth of elements is exponential, in real scenarios the number of available legs m is limited by thresholds that define the maximum SINR difference from the best leg, while L is usually limited by the number of UE antennas. Therefore, both m and L are, in practice, usually small numbers.

In our numerical evaluation, we execute the heuristic algorithm with $\eta \in \{5\%, 10\%\}$ as well as without fixing any value for η (label "*Heuristic w/o* η"). Disabling η corresponds to selecting the best combination of legs for each UE independently on its cardinality and number of UEs that have been already configured in Multi Connectivity.

Algorithm 2: PDD Heuristic Algorithm

Data: L: maximum number of legs,
η: maximum percentage of users in MC.
Result: Value of variables $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$

1 $\forall u \in \mathcal{U}, \forall l \in \mathcal{L}_u$ $y_{ul} \leftarrow 1$, $z_u \leftarrow 1$;
$\forall u \in \mathcal{U}, \forall l \in \mathcal{L}_u$ compute SINR γ_{ul};
$\forall u \in \mathcal{U}$ compute $\Gamma_u \leftarrow \prod_{l \in \mathcal{L}_u} \gamma_{ul}$;
$\mathcal{U}' \leftarrow Sort\left(\mathcal{U}, \frac{D_u}{\Gamma_u}, \text{ascending}\right)$;
2 **forall** $u \in \mathcal{U}'$ **do**
  Generate tuples of size up to L with different legs in $\mathcal{L}_u$: $\mathcal{T}_u = \{\mathcal{M}_u \subseteq \mathcal{L}_u : |\mathcal{M}_u| \leq L\}$;
  $\mathcal{B}_u = \emptyset$;
  forall $\mathcal{M}_u \in \mathcal{T}_u$ **do**
3 Feasible $\leftarrow$ false;
    forall $l \in \mathcal{M}_u$ **do**
      Compute BLER target $\phi_{ul} \leftarrow \frac{\Phi_u}{\gamma_{ul}}$;
      Select highest MCS $m \in \mathcal{M}$ such that $\phi_{ms} \geq \phi_{ul}$;
      if $R_m \leq D_u \wedge$ (22) *holds true* **then**
        $y_{ul} \leftarrow 0$, $x_{ulms} \leftarrow 0$;
        Feasible $\leftarrow$ false;
      else
        $y_{ul} \leftarrow 1$, $x_{ulms} \leftarrow 1$;
        Feasible $\leftarrow$ true;
      end
    end
4 **if** *Feasible* $\wedge\ \mathcal{M}_u \succ \mathcal{B}_u$ **then**
      if $\sum_{u \in \mathcal{U}: \{\sum_{l \in \mathcal{M}_u} y_{ul} > 1\}} z_u \leq \eta |\mathcal{U}|$ **then**
        if $|\mathcal{M}_u| > 1$ **then**
          $\forall l \in \mathcal{B}_u, y_{ul} \leftarrow 0$;
          $z_u \leftarrow 1$, $\forall l \in \mathcal{M}_u, y_{ul} \leftarrow 1$;
          $\mathcal{B}_u \leftarrow \mathcal{M}_u$;
        end
      else
        if $|\mathcal{M}_u| = 1$ **then**
          $\forall l \in \mathcal{B}_u, y_{ul} \leftarrow 0$;
          $z_u \leftarrow 1$, $\forall l \in \mathcal{M}_u, y_{ul} \leftarrow 1$;
          $\mathcal{B}_u \leftarrow \mathcal{M}_u$;
        end
      end
    else
      Reset all variables of u;
    end
  end
5 $\forall u \in \mathcal{U}, \forall l \in \mathcal{L}_u$ recompute SINR γ_{ul} with $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$;
  $\mathcal{U}' \leftarrow \mathcal{U}' \setminus \{u\}$;
  $\mathcal{U}' \leftarrow Sort\left(\mathcal{U}, \frac{D_u}{\Gamma_u}, \text{ascending}\right)$;
end

V. Numerical results

In this section, we illustrate the numerical results obtained solving the PSD and PDD problems in realistic network deployments. We first describe the evaluation methodology. Then we discuss the performance results of our proposed solutions.

A. Evaluation Methodology

In our numerical evaluation, we consider the 3GPP Urban Macro deployment scenario [6] depicted in Figure 2. The network is composed of 7 macro base stations with 3 sectors each, represented as black circles and colored hexagons, respectively. The inter-site distance between adjacent macro base stations is equal to 150 meters. Small cells, which are depicted as black crosses in the figure, are randomly placed in each sector according to a uniform distribution. The number of small cells per sector varies from 1 to 9. The transmission power, carrier frequency and system bandwidth are set to 43 dBm, 4 GHz, and 20 MHz for macro cells and 26 dBm, 6 GHz, and 20 MHz for small cells, respectively. For both layers, we consider an FDD system with all bandwidth allocated for downlink transmissions.

For each instance, we generate a random number of UEs in the $[2, 10]$ range and we randomly place the UEs in each sector in order to get evenly loaded sectors. In Figure 2, UEs are represented as colored squared dots with a color corresponding to the sector of the macro base station they belong to. Each

TABLE III: Parameters of the network deployment and traffic.

Parameter	**Description**
Deployment	Dense Urban
Number of macro cells	21 (7 gNBs with 3 sectors)
Number of small cells	$[1, 9]$ cells per sector
Number of UEs Total Number of UEs	$[2, 10]$ UEs per sector $[42, 210]$
Inter-site Distance	150 m
Channel model	Urban Macro (UMa)
Carrier frequency (macro)	4 GHz
Carrier frequency (small)	6 GHz
System bandwidth	20 MHz
Subcarrier spacing	15kHz
Number of symbols per slot	14
Number of subcarriers per PRB	12
Macro-cell power	43 dBm per 20 MHz
Small-cell power	26 dBm per 20 MHz
UE power	23 dBm
BLER target	$10^{[-5,-2]}$ or $10^{[-8,-5]}$
Highest MCS	64QAM
User data rate: low traffic regime and high traffic regime	$[0.1, 0.5]$ Mbit/s $[5, 9]$ Mbit/s
Power in active mode	300 power units

UE u is characterized by a traffic demand with data rate D_u and a BLER target (Φ_u) uniformly extracted at random in different ranges to take into account different QoS requirements. Specifically, user demands are drawn between $[0.1, 0.5]$ Mbit/s (Low Traffic regime) and $[5, 9]$ Mbit/s (High Traffic regime), whereas the BLER target is extracted from either $10^{[-5,-2]}$ or $10^{[-8,-5]}$ as specified in [33]. As indicated in Section III-C, we fix $\alpha = \frac{1}{|\mathcal{L}||\mathcal{U}|+1}$ to impose the maximization of the number of served users and the minimization of the used resources as primary and secondary objectives, respectively. Table III illustrates the parameters considered for the network deployment and user traffic in our numerical evaluation.

For each scenario, we consider 15 random extractions, and

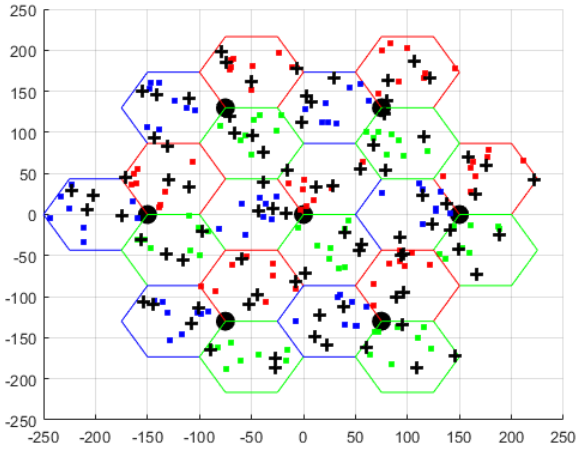


Fig. 2: Urban Macro deployment with Macro and Small Cells: black circles represent the 7 Macro Base Stations (Macro Cells are illustrated as hexagons), while Small Cells are depicted as black crosses. UEs are represented as colored squared dots (the color corresponds to the sector of the assigned Macro Cell).

we measure the average number of UEs accepted in the system, the number of activated legs, the BLER gap, the served rate and the time to compute the solution.

All numerical results have been obtained on a server equipped with an Intel(R) Xeon(R) E5-2640 v4 CPU@2.40GHz and 126 GB of RAM. Instances of network deployments and their corresponding parameters for the PSD and PDD problems have been generated using Matlab according to the system model described in Section III. The optimal solution has been computed using the CPLEX solver, using default values for its parameters and the time limit has been set to 3600 s, whereas the heuristic detailed in Algorithm 2 has been implemented in Matlab using the Parallel Computing Toolbox.

We first consider the PSD problem (Section V-B), providing a comparison of our proposed decomposition approach (*DA*) with a Single Connectivity solution, where only one leg is activated for each UE (i.e., with the Master gNB), thus quantifying the benefits obtained implementing Multi Connectivity. Then, we discuss the results obtained with DA in larger topologies, analyzing numerically the impact of the number of legs that can be activated on the system performance. Finally, we study the impact of the traffic offered to the network by each UE as well as of the number of installed small cells.

We then focus on the PDD problem (Section V-C), analyzing also in such case the effectiveness of the proposed heuristic (greedy algorithm) in providing good solutions, in terms of number of accepted users, computing time and served rate, and further studying how many legs are activated in average for satisfying the desired BLER target of each connection, along with the reliability (BLER) gap. This latter has been measured as the average across all users of the difference between the BLER seen by the user and its BLER target (i.e., the mean value of the slack variable of constraints (21)).

Note that in the numerical evaluation of the PSD problem we only consider eMBB users with BLER target fixed to 10%, whereas for the PDD problem we have considered both eMBB and URLLC users varying the BLER target as specified in the description below.

B. PSD Problem

1) Comparison with Single Connectivity: We first measure the benefit of activating and using multiple legs to carry user traffic. To this aim, we modified model (2)-(12) forcing UEs to be connected to the best available gNB using a single leg. Specifically, this is obtained solving problem (2)-(12) with objective function $\sum_{u\in\mathcal{U}} z_u$ and the following additional constraint:

$$\sum_{l\in\mathcal{L}_u} y_{ul} \leq 1 \qquad \forall u \in \mathcal{U}.$$

We evaluate a network composed of 5 to 7 small cells per sector, and 5 to 7 users per sector (i.e., with a total of 105 to 147 UEs, and 105 to 154 gNBs, respectively) The demand of each UE is uniformly extracted at random in the interval $[3, 8]$

Single Connectivity	5 SCs	7 SCs
5 UEs	54.3	58.8
7 UEs	67.6	73.9

TABLE IV: Average number of users accepted in the system for the Single Connectivity case.

Mbit/s. The results comparing single and dual connectivity scenarios are shown in Tables IV and V, respectively.

Dual Connectivity	5 SCs	7 SCs
5 UEs	65.5	73.3
7 UEs	87.0	98.3

TABLE V: Average number of users accepted in the system for the Dual Connectivity case (2 legs available per UE).

In these scenarios, the number of served users in Dual Connectivity is between 20.6% (when UEs=Small Cells= 5 per sector) and 33% (for UEs=Small Cells= 7 per sector) higher than in Single Connectivity. The gain is larger when the number of interfering UEs in the network increases, as well as when a large number of available connections to gNBs exists (either master or small cells).

Having the possibility to activate more than 2 legs has, in this scenario, a limited impact on the number of served users. In the same scenario illustrated above, we observed that having up to 4 legs available per user permits to further increase the number of accepted users, in average of less than 8%, at the price of increasing the solving time of a factor larger than 2.5.

2) Analysis of multiple cell choices: The average number of accepted users and selected legs as a function of the number of UEs per sector are illustrated in Figures 3 and 4. In the figures, 2 and 5 small cells per sector are deployed and users can simultaneously activate either 2 or 5 legs. Each user is characterized by a demand with data rate uniformly extracted at random in the $[1;5]$ Mbit/s range. Curve label $LX.SCY$ indicates that each UE can activate and use at most *X Legs* out of *Y* available *Small Cells*.

By comparing the two curves $L2.SC5$ and $L5.SC2$ in Figure 3 (in the middle), we can conclude that the mobile operator can obtain the same performance by either increasing the network coverage (i.e., increasing the number of small cells per sector from 2 to 5) or activating more legs for each user (i.e., increasing the active legs from 2 to 5). Therefore, deploying more small cells becomes crucial to increase system throughput especially in dense urban scenarios. This is confirmed by the results of the two $L2.SC2$ and $L2.SC5$ curves in Figure 4 (at the bottom). On average, having a larger number of Small Cells deployed in the network results in the activation of more legs per user, which permits to serve more users. For example, when we have 10 UEs per cell, 2 Small Cells per sector and each UE can select up to 5 legs, the number of served users increases by 33% (134.1 versus 100.9), while the number of active legs grows by 13% (1.31 with respect to 1.16).

In Figure 3 we also illustrated for every point the 95% confidence intervals obtained in our numerical analysis, which are very small. In all the results we measured in the evaluation campaign, indeed, the width of the 95% confidence interval was, for each point, always smaller than 5% the average value of the point itself (in the worst case), and in several cases was even lower than 3%. For this reason and for the sake of clarity, we did not report such confidence intervals in the following figures, to avoid cluttering them, since sometimes they contain multiple lines.

3) Analysis of demand and number of small cells: In this performance evaluation, we increased the traffic load of each UE by varying it in the interval $[1;9]$ Mbit/s. Specifically, we evaluate five traffic scenarios: D1, where each UE has a data rate randomly drawn from a uniform distribution in the range $[1;5]$ Mbit/s, D2 ($[2;6]$ Mbit/s), D3 ($[3;7]$ Mbit/s), D4 ($[4;8]$ Mbit/s), and D5 ($[5;9]$ Mbit/s). Figures 5,6,8 illustrate the corresponding results, where the number of small cells in each of the 21 sectors is increased from 1 to 7, while the number of UEs per sector is set to 5, which corresponds to a total of 105 UEs.

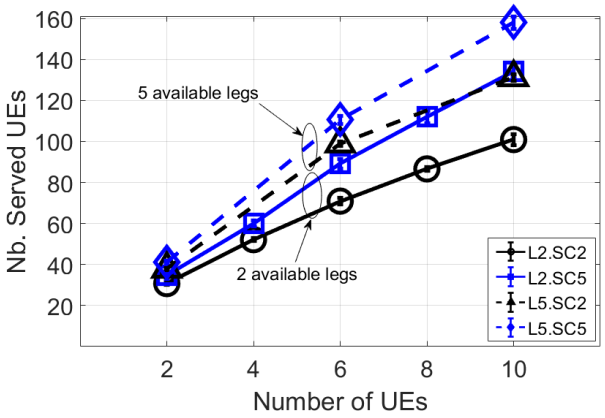


Fig. 3: Average number of served UEs as a function of (i) the number of UEs per sector (x axis), (ii) the number of small cells per sector (2,5) and (iii) the number of legs that can be activated (2,5); 95% confidence intervals are also displayed.

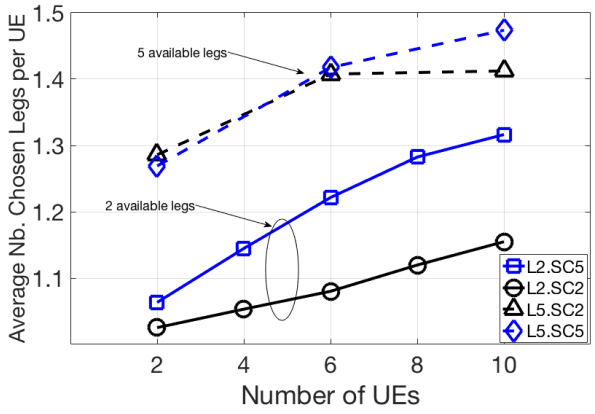


Fig. 4: Average number of selected legs as a function of (i) the number of UEs per sector (x axis), (ii) the number of small cells per sector (2,5) and (iii) the number of legs that can be activated (2,5).

The average *number of accepted users* (Figure 5) depends, as expected, on the demand of each user, and on the number of small cells deployed in each sector. When 3 small cells are installed per sector, our model accepts on average 71 users for the lowest demand (scenario D1 with $[1; 5]$ Mbit/s), 64.7 for the medium one (scenario D3 with $[3; 7]$ Mbit/s) and 61.5 for the highest demand (scenario D5 with $[5; 9]$ Mbit/s). If we focus on a given traffic load scenario, the number of accepted users increases from 62.6 to 83.27 and from 48.9 to 76.4 for the D1 and D5 scenarios, respectively.

The average *number of legs* simultaneously enabled by each UE increases with both the traffic load and the number of small cells deployed in each sector as shown in Figure 6. The increase depends roughly equally on both parameters. In particular, increasing the number of small cells per sector from 1 to 7 yields an increase of active legs in the order of 12.2% and 18.7% for the scenario D1 and D5, respectively. We observe that a larger number of active legs corresponds to an increase in power consumption of the user devices as illustrated in Figure 7. This is due to the larger spectrum that users must monitor.

Figure 8 illustrates the average *computing time* to obtain a solution (in seconds). In all considered scenarios where at least 3 small cells per sector are deployed, the computing time is fairly low (at most up to 740 seconds, in average). In particular, increasing the number of small cells deployed in each sector results in general in a higher number of potential connections that can be activated to serve the user traffic. In this case, our heuristics can find a solution quickly since the SINR constraints are not the limiting factor of the problem instance. Note that solving the PDD problem is slightly more difficult, in terms of computing time, in the medium scenario we considered, D3, than for the higher demand scenario D5, even though such difference tends to become much less evident when the number of small cells per sector increases. This is due to the fact that, when the traffic offered by each user becomes large, several users are not admitted in the network and the solver can reach the optimum significantly faster. This effect is more evident when the number of macro/small cells deployed in the network (able to accept and handle users' demands) is smaller, e.g., equal to 1. We observe that real 5G network scenarios will deploy a significant number of small cells (e.g., at least 4 per sector [34]). Therefore, our proposed methods represents a practical solution for planning 5G and 5G-Advanced cellular networks.

We would like to note that in this paper we focus on solving a planning problem that is not time-critical, and it is also for this reason that we do not explicitly model the scheduling policies. A planning problem like the ones we consider in this paper can be solved offline for multiple deployment instances and the corresponding solutions can be stored in a database to be used during the operation of the network. Therefore, the different runs that we consider for each network scenario can be interpreted as possible placement scenarios of users (e.g., the most frequent ones) and the corresponding solutions can be stored in a database. During the network operation, the management system needs only to check which users placement is the closest to the ones stored in the database and apply the corresponding configuration. The notion of proximity between actual and stored users placements can be modeled considering different norms and system metrics as norms input. Typical example of norms are L-norms while path-loss or user position can be used as norm inputs. We observe that the way of operating and controlling the network in real-time is out of the scope of this paper.

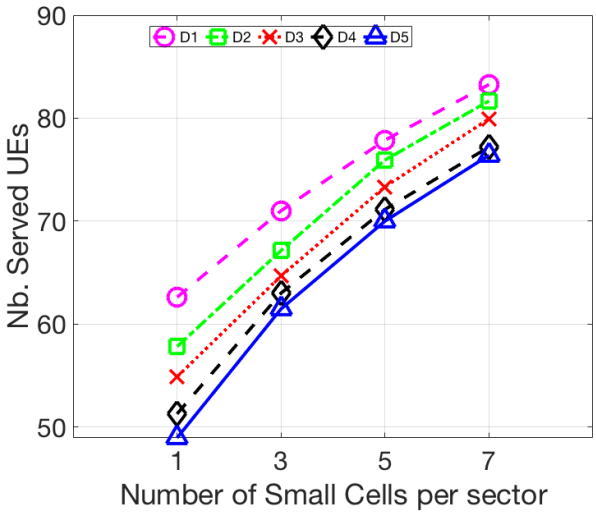


Fig. 5: Average number of served users as a function of the number of small cells per sector and UEs' demand. We considered 5 scenarios with increasing demand, D1 - D5.

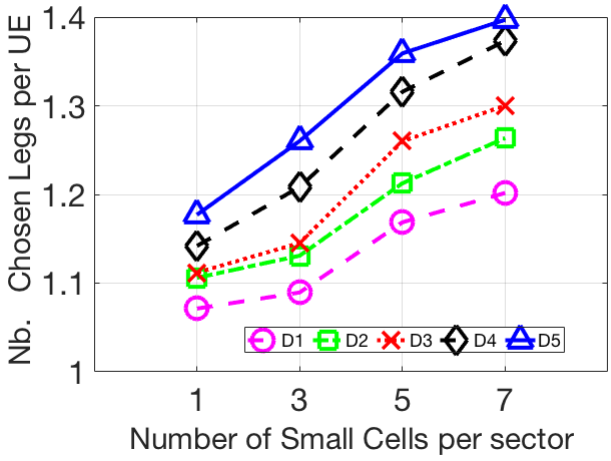


Fig. 6: Average number of chosen legs per UE as a function of the number of small cells per sector and UEs' demand.

C. PDD Problem

1) Comparison with Single Connectivity: Also in this case we first quantify the benefit that can be obtained by activating more than one leg for each user, implementing a variation of

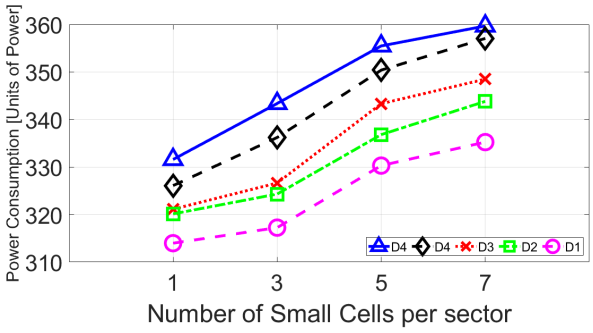


Fig. 7: Average power consumption per UE as a function of the number of small cells per sector and UEs' demand. Note that 3GPP defines no mapping of [units of power] to standard units of power like Watt.

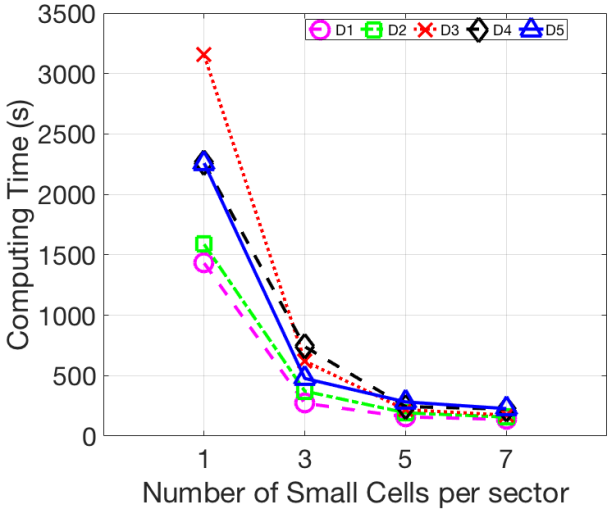


Fig. 8: Average computing time (seconds) as a function of the number of small cells per sector and UEs' demand.

model (13)-(23) that forces UEs to be connected with a single leg to the best available gNB.

Single Connectivity	5 SCs	9 SCs
6 UEs	67.7	78.2
8 UEs	81.5	98.7
10 UEs	100.7	116.2

TABLE VI: Average number of accepted users for the Single Connectivity case (with variable numbers of UEs and Small Cells).

Multi Connectivity	5 SCs	9 SCs
6 UEs	86.3	100.3
8 UEs	107.2	123.3
10 UEs	127.3	148.9

TABLE VII: Average number of accepted users for the Multi Connectivity case (multiple legs available per UEs, with variable numbers of UEs and Small Cells).

It can be observed that the number of served users when each UE can activate multiple legs as illustrated in Table VII is 25% to 31.5% higher than in the Single Connectivity scenario (Table VI). The gain is hence consistent throughout all network instances.

2) Number of accepted Users and Served Rate: Figure 9 illustrates the average number of served users as a function of the number of small cells and UEs per sector, which range from 1 to 9 and from 2 to 10, respectively. It can be observed that the heuristic (*Heuristic w/o* η, dashed lines) is practically overlapping with the optimum obtained solving the PDD model (solid lines) when the number of UEs is small (2 and 4 per sector). For higher UE values, the heuristic provides solutions that are at most 8.1% and 8.7% lower than the optimum, in the worst case (i.e., when the number of small cells per sector is equal to 9), when the UEs are 6 and 8 per sector. When 10 UEs per sector are deployed in the network, such gap is at most equal to 20%. At the same time, as we will show later, the heuristic is able to compute such good solutions within, at most, hundreds of seconds in the largest network scenarios, hence more than one order of magnitude lower than the time limit imposed to CPLEX.

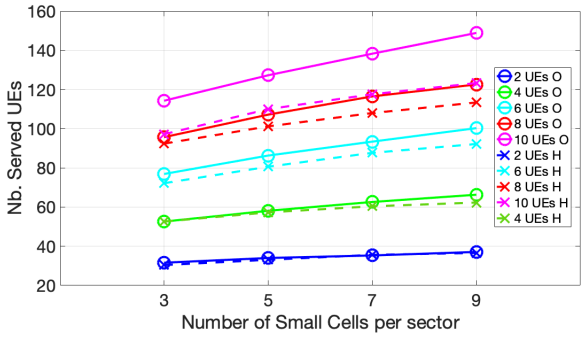


Fig. 9: Average number of served users as a function of the number of Small Cells and UEs per sector, obtained with both the optimum computed solving the PDD model (O, solid lines) and the proposed Heuristic w/o η (H, dashed lines).

Figure 10 illustrates the same performance figure (average number of accepted users), comparing the results obtained using the original data rate for the demands as well as the BLER target (i.e., $10^{[-2;-5]}$), and two scenarios where we reduce the traffic data rate by a factor of 10 and increase the BLER target in the range $10^{[-5;-8]}$. In Figure 10, we label the curves corresponding to these two scenarios as Low Traffic (LT) and High BLER (HB), respectively. It can be observed that in the considered network scenario, setting a higher BLER target has a limited impact on the number of accepted users (the HB curves are practically overlapping to the solid ones, especially for a small number of UEs per sector, and only slightly lower for higher UE values). This is due to the fact that in such scenarios it is the channel capacity to limit the performance of the system, and not reliability constraints (i.e., the BLER target set for each user). For the same reason, a lower traffic demand (LT dashed lines) allows the Mobile network operator to accept more users (up to 19.7% in the

best case).

The average total served rate (expressed in Mbps) and the average utilization of radio resources (expressed in percentage of available PRBs) measured in the same scenario are illustrated in Figures 11 and 12, respectively. The served rate follows the same trend as in Figure 10, whereas the utilization of radio resources decreases as the number of Small Cells deployed in the system increases. As the number of deployed Small Cells increases, the average distance and the path loss between a user and its nearby SCs decreases. Therefore, the user can be served with faster MCS, thus reducing its utilization of PRBs at equal traffic. We further observe that when the number of users increases, the system becomes limited by the interference, which reduces the SINR and makes more challenging fulfilling the target error probability. Therefore, in highly dense and highly loaded scenarios the system may become limited by error probability constraints rather than capacity constraints when QoS constraints must be guaranteed for all users and at any time. Note however that in other scenarios with a mix of users and different QoS metrics, capacity may still be the main limiting factor for the number of accepted users.

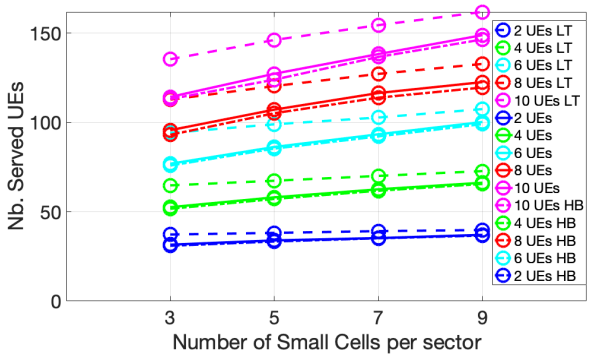


Fig. 10: Average number of served users as a function of the number of Small Cells and UEs per sector, obtained solving the PDD model at the optimum for (1) the original demand (uniformly distributed between 2 and 5 Mbps per user, solid lines), (2) a Low Traffic (LT) regime with ten times lower demands (between 0.1 and 0.5 Mbps per user, dashed lines) and (3) a Higher BLER target (dash-dotted lines, HB) value, equal to $10^{[-5,-8]}$ (for the two other lines this was set to the default value we considered in our numerical evaluation, i.e. $10^{[-2,-5]}$).

Finally, Figure 13 compares the optimal and heuristic solutions in terms of number of served users in the largest deployment scenario with 10 UEs per sector. The heuristic is configured both disabling the threshold η and with $\eta = \{5, 10\}\%$. The optimum is depicted using a solid line while the three heuristic solutions are illustrated using dashed lines. When the fraction of users that are configured with Multi Connectivity increases, up to 10%, the total number of users that can be accepted in the system decreases, since more resources are dedicated to a single user which enables more than one leg. Specifically, such decrease is in the order of

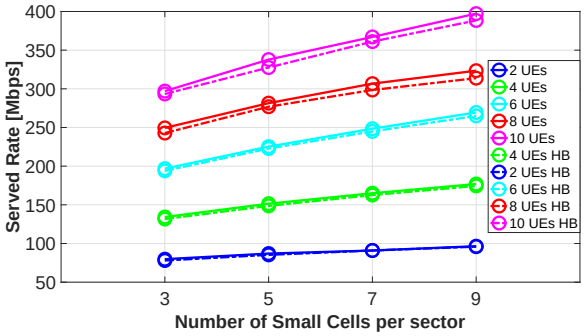


Fig. 11: Average Served Rate (total), in Mbps, as a function of the number of Small Cells per sector, obtained solving the PDD model at the optimum (solid line, BLER target equal to $10^{[-2,-5]}$), and the same model imposing a higher BLER target, equal to $10^{[-5,-8]}$ (dash-dotted line, HB).

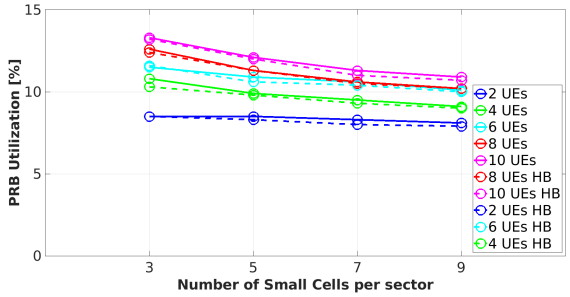


Fig. 12: Average utilization of radio resources (PRBs) as a function of the number of Small Cells per sector, obtained solving the PDD model at the optimum (solid line, BLER target equal to $10^{[-2,-5]}$), and the same model imposing a higher BLER target, equal to $10^{[-5,-8]}$ (dash-dotted line, HB).

21.4% (in average) when $\eta = 5\%$ and up to 35.8% when $\eta = 10\%$. This permits to provide a first quantification of the trade-off that exists between serving more users and providing higher quality of service (in terms of reliability, for example, as we will show in the following when commenting the results related to the BLER gap with respect to the requirements, as well as in terms of the number of activated legs).

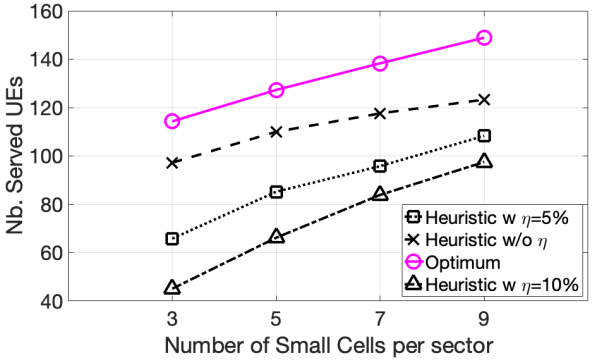


Fig. 13: Average number of served users as a function of the number of Small Cells per sectors, for 10 UEs per sector, obtained solving the PDD model at the optimum (solid line) or the Heuristic with different threshold values for the percentage of users configured with MC (*Heuristic w/o* η and *w* $\eta = 5, 10\%$).

3) Analysis of the Reliability (BLER) Gap: We now study the Reliability Gap in different scenarios. Figure 14 illustrates the average BLER Gap per user as a function of the number of small cells per sector, obtained (1) solving the PDD model at the optimum (solid line, BLER target equal to $10^{[-2,-5]}$), (2) solving the same model at the optimum with a Low Traffic regime, with 1/10th of the demand of the previous scenario (dashed line), (3) using the Heuristic without threshold η for a BLER target equal to $10^{[-2,-5]}$ (dotted line) and finally (4) using the same heuristic with a more stringent BLER target, equal to $10^{[-5,-8]}$ (dash-dotted line).

Solving the model at the optimum permits to obtain solutions in which the served users have a larger gap in terms of assigned resources with respect to their target BLER, and this is reflected in the two higher curves. Of course, when demands are lower, such gap also increases especially when more Small Cells are available (see the highest curve in the figure, LT). As for the heuristics (the two lower curves), they provide solutions with (much) tighter gaps, and having users with more stringent BLER requirements (BLER target equal to $10^{[-5,-8]}$) leads to gaps that are, in average, 31.7% lower than those computed by the same heuristics with BLER target equal to $10^{[-2,-5]}$.

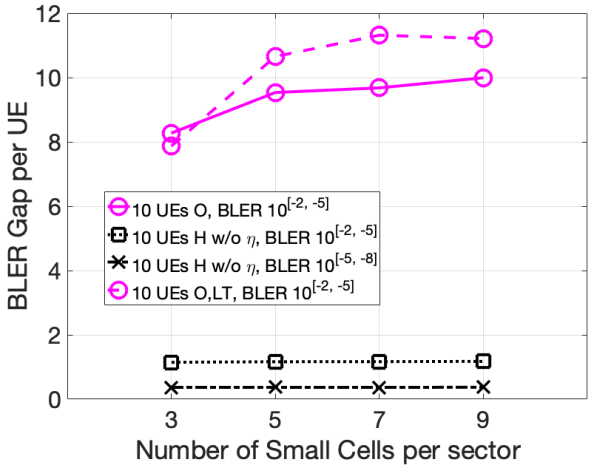


Fig. 14: Average BLER Gap per UE as a function of the number of small cells per sector, obtained solving the PDD model at the optimum (solid line, BLER target equal to $10^{[-2,-5]}$), the same model with a Low Traffic regime (dashed line, O, LT), the Heuristic without threshold η for both a BLER target equal to $10^{[-2,-5]}$ (dotted line) and $10^{[-5,-8]}$ (dash-dotted line).

4) Number of chosen legs: The number of legs activated by each user is an indicator of both the amount of resources consumed in the network and the reliability achieved by the solution (i.e., the higher the number of used legs, the higher the consumed resources and reliability). Figure 15 illustrates the average number of legs chosen by the model (the solid line) and the greedy heuristics (without η and with $\eta = 5, 10\%$). It can be observed that the curve related to the optimum is between the one where 10% of the users activate Multi Connectivity and the one with a 5% threshold. In the considered scenarios, in practice, the best tradeoff between resource consumption and reliability can be obtained in such a range of fraction of the users. The gap in the number of activated legs for these two cases is up to 25% (about 1.5 vs 1.2, for SC=3), while the same number is much lower when we do not impose any threshold η to the heuristic. Figure 16 shows the average power consumed by a user device as a function of the number of SCs per sector in the system. Similar to the PSD problem, the power consumption of a UE increases with the number of active legs. Note that the heuristic tends to activate less legs than the optimum when the number of SCs increases, since a user has a larger probability of being served by a single SC. This also drives down the power consumed by a user.

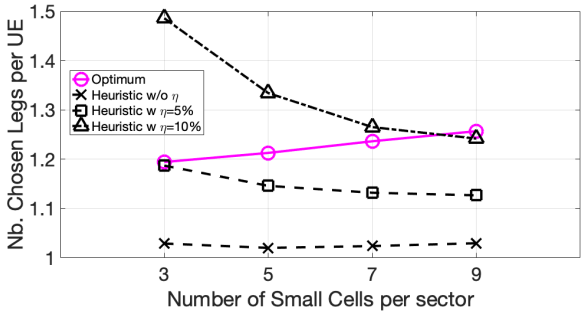


Fig. 15: Average number of chosen legs per user as a function of the number of Small Cells per sectors, for 10 UEs per sector, obtained solving the PDD model at the optimum (solid line), a Low Traffic regime (LT) with ten times lower demands (between 0.1 and 0.5 Mbps per user, dashed lines) or the Heuristic with different threshold values for the percentage of users configured with MC (without η and with $\eta = 5, 10\%$).

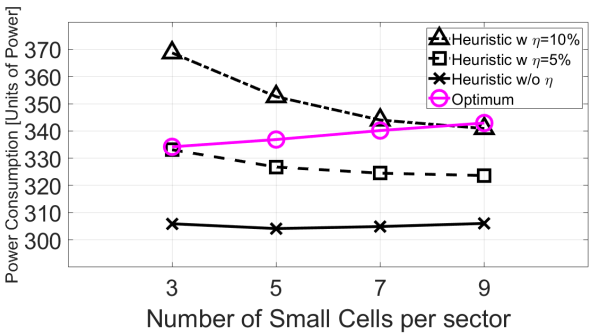


Fig. 16: Average power consumption per user as a function of the number of Small Cells per sectors, for 10 UEs per sector, obtained solving the PDD model at the optimum (solid line), a Low Traffic regime (LT) with ten times lower demands (between 0.1 and 0.5 Mbps per user, dashed lines) or the Heuristic with different threshold values for the percentage of users configured with MC (without η and with $\eta = 5, 10\%$).

5) Computing Time: The average computing time is illustrated in Figure 17 as a function of the number of Small Cells per sectors, for a selected number of UEs (large, medium and small, i.e. 10, 6 and 2 UEs per sector), necessary to solve at the optimum the PDD problem considering also a Low Traffic Regime.

The computing time increases as expected with the network instance size, specifically when the number of UEs increases.

Solving the problem at the optimum requires a consistent amount of time. In practice, for the largest instances (10 UEs per sector), the time limit of 3600s imposed to CPLEX is always reached, so that the result is not guaranteed to be optimal. Considering a Low Traffic regime permits to obtain solutions in a shorter amount of time, since in such scenario all constraints are loose and CPLEX therefore solves an unconstrained optimization problem.

Finally, we observe that our greedy heuristic is able to compute solutions within, at most, hundreds of seconds in the largest network scenarios, hence more than 10 times lower than the time limit imposed to CPLEX. The computing time of the heuristic algorithm increases with the number of small cells since the list of combinations of legs for each UE increases. Indeed, a higher density of small cells results in a larger set of legs that the UE can simultaneously activate. This trend is not visible in the optimal solution since we fix the maximum computing time for CPLEX.

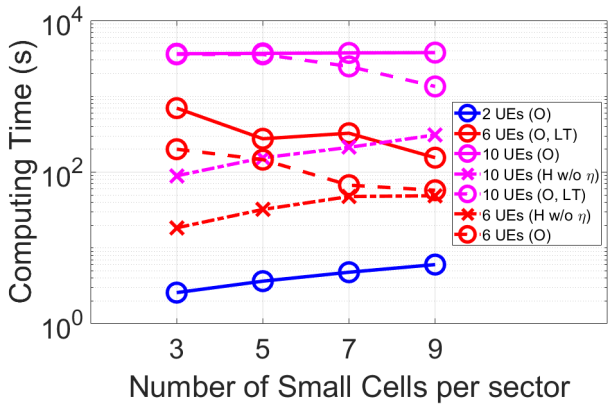


Fig. 17: Average Computing Time (logarithmic scale) as a function of the number of Small Cells per sector, for 2, 6 and 10 UEs per sector, obtained solving the PDD model at the optimum (solid lines, O), the Optimum for a Low Traffic regime (O, LT) with ten times lower demands (between 0.1 and 0.5 Mbps per user, dashed lines) and the Heuristic w/o threshold η (dashed-dotted lines, H). For UE=2 only the optimum with full traffic demands is shown since computing times are very small. The time limit was set to 3600s for CPLEX.

VI. Conclusions

In this paper we addressed the PDCP Split-Bearer and the Duplication Decision problems in 5G networks with Multi-Connectivity capabilities. To solve these problems we developed an optimization framework that decides which users to admit in the system, whether to activate multiple connections to satisfy user requirements and how to allocate radio resources.

More in detail, for both PSD and PDD problems, we proposed an optimization model based on mathematical programming that achieves the optimal solution and heuristic algorithms to speed up the computation of a solution for real-size network scenarios.

We performed a thorough numerical evaluation, considering realistic mobile network deployments and traffic scenarios. For PSD, we observed an increase in the average number of accepted users in the network up to 33% with respect to a baseline approach that implements Single Connectivity, while for PDD such gain was up to 31.5%. The evaluation further captures and quantifies the trade-off between serving a large number of users and providing the necessary radio resources needed to satisfy the throughput and reliability requirements.

Numerical results show that deploying denser networks and enabling multi-connectivity permit to increase user reliability and system throughput when radio resources are wisely allocated to limit interference.

References

[1] Ijaz Ahmad, Tanesh Kumar, Madhusanka Liyanage, Jude Okwuibe, Mika Ylianttila, and Andrei Gurtov. Overview of 5G security challenges and solutions. *IEEE Commun. Standards Magazine*, 2(1):36–43, 2018.

[2] Younsun Kim, Youngbum Kim, Jinyoung Oh, Hyoungju Ji, Jeongho Yeo, Seunghoon Choi, Hyunseok Ryu, Hoondong Noh, Taehyoung Kim, Feifei Sun, et al. New radio (NR) and its evolution toward 5G-advanced. *IEEE Wireless Communications*, 26(3):2–7, 2019.

[3] Amitabha Ghosh, Andreas Maeder, Matthew Baker, and Devaki Chandramouli. 5G evolution: A view on 5G cellular technology beyond 3GPP release 15. *IEEE Access*, 7:127639–127651, 2019.

[4] Adnan Aijaz. Packet Duplication in Dual Connectivity Enabled 5G Wireless Networks: Overview and Challenges. *IEEE Communications Standards Magazine*, 3(3):20–28, 2019.

[5] Jocelyne Elias, Fabio Martignon, and Stefano Paris. Optimal Split Bearer Control and Resource Allocation for Multi-Connectivity in 5G New Radio. In *2021 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit)*, pages 187–192. IEEE, 2021.

[6] 3GPP tech. report (TR) 36.872. Small cell enhancements for E-UTRA and E-UTRAN - physical layer aspects. Technical report, 3GPP, 2013.

[7] Jose F. Monserrat, Faiza Bouchmal, David Martin-Sacristan, and Oscar Carrasco. Multi-Radio Dual Connectivity for 5G Small Cells Interworking. *IEEE Communications Standards Magazine*, 4(3):30–36, 2020.

[8] A. Ravanshid, P. Rost, D.S. Michalopoulos, V.V. Phan, H. Bakker, D. Aziz, S. Tayade, H.D. Schotten, S. Wong, and O. Holland. Multi-connectivity functional architectures in 5G. In *IEEE International Conference on Communications Workshops (ICC)*, pages 1–6, 2016.

[9] Hongtao Zhang and Wanqing Huang. Tractable mobility model for multi-connectivity in 5G user-centric ultra-dense networks. *IEEE Access*, 6:43100–43112, 2018.

[10] Margarita Gapeyenko, Vitaly Petrov, Dmitri Moltchanov, Mustafa Riza Akdeniz, Sergey Andreev, Nageen Himayat, and Yevgeni Koucheryavy. On the Degree of Multi-Connectivity in 5G Millimeter-Wave Cellular Urban Deployments. *IEEE Transactions on Vehicular Technology*, 68(2):1973–1978, 2019.

[11] Waqar Anwar, Atul Kumar, Norman Franchi, and Gerhard Fettweis. Physical layer abstraction for multi-connectivity communications: Modeling and analysis. *IEEE Trans. on Wireless Communications*, 2021.

[12] Yang Yang, Xinyi Deng, Dazhong He, Yanan You, and Ruoning Song. Machine Learning Inspired Codeword Selection For Dual Connectivity in 5G User-Centric Ultra-Dense Networks. *IEEE Transactions on Vehicular Technology*, 68(8):8284–8288, 2019.

[13] Lingxia Wang, Chungang Yang, and Rose Qingyang Hu. Autonomous traffic offloading in heterogeneous ultra-dense networks using machine learning. *IEEE Wireless Communications*, 26(4):102–109, 2019.

[14] Jani Saloranta, Giuseppe Destino, Antti Tölli, and Henk Wymeersch. Novel solution for multi-connectivity 5G-mmW positioning. In *52nd IEEE Asilomar Conference on Signals, Systems, and Computers*, pages 537–540, 2018.

[15] Elizaveta Rastorgueva-Foi, Olga Galinina, Mário Costa, Mike Koivisto, Jukka Talvitie, Sergey Andreev, and Mikko Valkama. Networking and Positioning Co-Design in Multi-Connectivity Industrial mmW Systems. *IEEE Trans. on Vehicular Technology*, 69(12):15842–15856, 2020.

[16] J. Rao and S. Vrzic. Packet Duplication for URLLC in 5G: Architectural Enhancements and Performance Analysis. *IEEE Network*, 32(2):32–40, March 2018.

[17] N.H. Mahmood, A. Karimi, G. Berardinelli, K.I. Pedersen, and D. Laselva. On the resource utilization of multi-connectivity transmission for URLLC services in 5G New Radio. In *IEEE Wireless Communications and Networking Conference Workshop (WCNCW)*, pages 1–6, 2019.

[18] Jaya Rao and Sophie Vrzic. Packet duplication for URLLC in 5G dual connectivity architecture. In *IEEE Wireless Communications and Networking Conference (WCNC)*, 2018.

[19] Nurul Huda Mahmood and Hirley Alves. Dynamic multi-connectivity activation for ultra-reliable and low-latency communication. In *IEEE International Symposium on Wireless Communication Systems (ISWCS)*, pages 112–116, 2019.

[20] Jihas Khan and Lillykutty Jacob. Link adaptation for multi-connectivity enabled 5g urllc: Challenges and solutions. In *2021 International Conference on COMmunication Systems NETworkS (COMSNETS)*, pages 148–152, 2021.

[21] Xinran Ba and Yafeng Wang. Load-aware cell select scheme for multi-connectivity in intra-frequency 5G ultra dense network. *IEEE Communications Letters*, 23(2):354–357, 2019.

[22] Konstantinos Alexandris, Chia-Yu Chang, Kostas Katsalis, Navid Nikaein, and Thrasyvoulos Spyropoulos. Utility-based resource allocation under multi-connectivity in evolved LTE. In *IEEE 86th Vehicular Technology Conference (VTC-Fall)*, pages 1–6, 2017.

[23] Marco Centenaro, Daniela Laselva, Jens Steiner, Klaus Pedersen, and Preben Mogensen. Resource-efficient dual connectivity for ultra-reliable low-latency communication. In *91st IEEE Vehicular Technology Conference (VTC2020-Spring)*, pages 1–5, 2020.

[24] Marco Centenaro, Daniela Laselva, Jens Steiner, Klaus Pedersen, and Preben Mogensen. System-Level Study of Data Duplication Enhancements for 5G Downlink URLLC. *IEEE Access*, 8:565–578, 2020.

[25] Roberto P Antonioli, Emanuel B Rodrigues, Diego A Sousa, Igor M Guerreiro, Carlos FM e Silva, and Francisco RP Cavalcanti. Adaptive bearer split control for 5G multi-RAT scenarios with dual connectivity. *Computer Networks*, 161:183–196, 2019.

[26] SungHoon Jung and SaeWoong Bahk. Online Control of Traffic Split and Distributed Cell Group State Decisions for Multi-connectivity. In *IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1–6. IEEE, 2020.

[27] SungHoon Jung and SaeWoong Bahk. Online Control of Traffic Split and Distributed Cell Group State Decisions for Multi-connectivity in 5G and Beyond. *IEEE Transactions on Vehicular Technology*, 2021.

[28] SungHoon Jung and SaeWoong Bahk. Real-time dual-link transmission control for minimizing power and link switching cost. *Computer Networks*, 200:108466, 2021.

[29] Prabodh Mishra, Snigdhaswin Kar, Vikas Bollapragada, and Kuang-Ching Wang. Multi-Connectivity using NR-DC for High Throughput and Ultra-reliable Low Latency Communication in 5G Networks. In *IEEE 4th 5G World Forum (5GWF)*, pages 36–40, 2021.

[30] Diomidis S. Michalopoulos and Volker Pauli. Data Duplication for High Reliability: A Protocol-Level Simulation Assessment. In *IEEE International Conference on Communications (ICC)*, pages 1–7, 2019.

[31] Yu-Ngok Ruyue Li, Mengzhu Chen, Jun Xu, Li Tian, and Kaibin Huang. Power saving techniques for 5g and beyond. *IEEE Access*, 8:108675–108690, 2020.

[32] Roch Guerin, Hamid Ahmadi, and Mahmoud Naghshineh. Equivalent capacity and its application to bandwidth allocation in high-speed networks. *IEEE Journal on selected areas in communications*, 9(7):968–981, 1991.

[33] 3GPP tech. specification (TS) 22.104. Service requirements for cyber-physical control applications in vertical domains. Technical report, 3GPP, 2021.

[34] Nurul Huda Mahmood, Melisa Lopez, Daniela Laselva, Klaus Pedersen, and Gilberto Berardinelli. Reliability Oriented Dual Connectivity for URLLC services in 5G New Radio. In *15th IEEE International Symposium on Wireless Communication Systems (ISWCS)*, 2018.