2025 36th International Teletraffic Congress (ITC 36)

Providing Throughput Guarantees in Multi-Domain Networks in 6G

Fidan Mehmeti
Chair of Communication Networks
Technical University of Munich, Germany
Email: fidan.mehmeti@tum.de

Wolfgang Kellerer
Chair of Communication Networks
Technical University of Munich, Germany
Email: wolfgang.kellerer@tum.de

***Abstract*—While the deployment of 5G networks is going on at a high pace, the research and industrial community have already started looking into the next generation of cellular networks, 6G. One of the main features envisioned in 6G are multi-domain networks, where both public networks (owned by cellular network operators) and private networks (owned by different users/institutions) will be deployed and would need to inter-operate in order to provide a satisfying level of service to the users. However, as these networks are operated by different entities, it is very challenging to provide end-to-end guarantees to a user whose data traverse multiple networks before reaching the destination, either in terms of the maximum latency, minimum throughput, or reliability. In this work, we focus on soft throughput guarantees. The approach we follow here is to use statistical knowledge from the activity of users or their data rate if the latter is constant, which can be obtained by the corresponding networks, in other network domains to determine the throughput range. We provide an analytical approach that determines this range, depending on how narrow the range needs to be. The evaluation is performed on input data from a publicly-available dataset. Results show that the soft guarantees lead to unchanged low-variability end-to-end throughput with more efficient resource utilization (an improvement of 18%) than to focus on strict rates and/or policies oblivious to other networks, such as Round-robin.**

***Index Terms*—6G, Campus networks, Performance guarantees, Throughput, Multi-domain networks.**

I. Introduction

5G networks have been introduced and are currently being deployed quickly to provide services that could not have been offered before, pertaining to one of the three service types: enhanced mobile broadband (eMBB) [1], ultra-reliable low-latency communications (URRLC) [2], and massive machine-type communications (mMTC) [3].

While significant performance improvements have been reported with 5G [4], [5], [6], still there are applications, such as holographic communications [7], for whose successful operation the available 5G resources do not suffice. These applications are very stringent in terms of the amount of network resources required (bandwidth-hungry applications), in terms of the time needed to deliver a packet (latency-sensitive applications), or in terms of reliability.

The other aspect that 5G networks do not cover are multi-domain networks [8], comprising several single-domain networks, known as *campus networks* [9]. These are private networks, including Radio Access Network (RAN) and Core Network (CN), not owned by the cellular operators, providing network access within a university, hospital, etc. Therefore, the research community backed up by industrial partners [8] already started working on 6G networks, planned to be fully operational by 2030 [10].

In line with the description of campus networks, 6G are planned to be heterolithic in terms of the network "owners" through which data traverse. Namely, the sender of the data can be in a network owned by a cellular network operator. The packets from the sender are transmitted via the wireless interface to the Base Station (BS) that serves it, from where the packets are forwarded to the CN. The receiver, on the other hand, may be within the coverage area of a campus network, not owned by the cellular operator, but by a private entity. Therefore, these packets from the CN of the "transmitter network" are forwarded through the Internet to the CN of the campus network, from where they are further forwarded to the corresponding BS of the campus network. Finally, the packets are delivered to the receiver. This is illustrated in Fig. 1.

Given the different operators managing the "transmitter" and "receiver" networks, providing any end-to-end performance guarantees, either in terms of throughput, latency, or reliability, in multi-domain networks can pose significant challenges. For example, if there is a minimum data rate at which an application needs to run, the transmitter experiencing given channel conditions would require a given amount of RAN resources to satisfy that data rate. The receiver, on the other end of the communication path, will most probably experience different channel conditions. Therefore, it will require a different amount of resources to satisfy the end-to-end rate requirement. However, it would be cumbersome for different entities operating different campus networks to exchange all the information on the channel conditions of all their users, so that the data rates on the transmitter and receiver network-side match. The other reason is that each campus network needs to maintain its privacy by disclosing only limited information to other campus networks.

The importance of matching data rates on both sides of the communication process stems from the fact that *the end-to-end throughput is determined by the lowest data rate in the cycle (path), i.e., by the bottleneck link.* Therefore, providing

978-3-948377-03-8/19/$31.00 ©2025 ITC

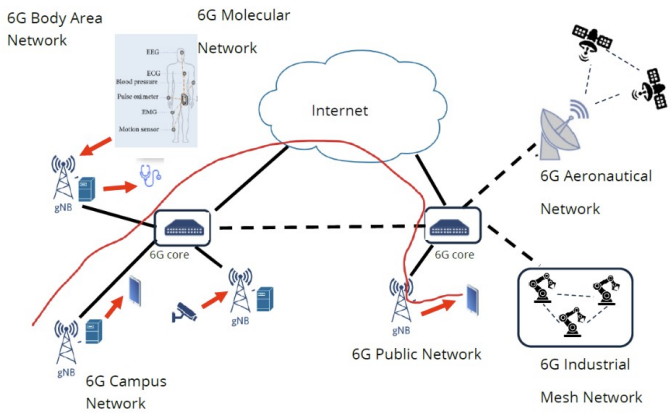

Fig. 1. Illustrating how the communication process in a 6G network comprising multiple campus networks and a classical (public) cellular network could look like. The red curve depicts a path example of the communication.

a data rate that is much higher on one of the sides leads to wasting resources as the throughput would be determined anyway by the data rate on the bottleneck link. Given the ever increasing number of these bandwidth-hungry applications and the finite resources, it would be highly inefficient to allow such a mismatch between the data rates on different ends of the communication cycle.

On the other hand, matching data rates is very challenging given the autonomous nature of the different network entities. To reconcile for these opposing requirements, in this paper we propose an approach which provides *soft* performance guarantees in terms of throughput. Instead of going for a strict rate, the data rates at both sides lie within a given interval exploiting limited information from other campus networks. This information can be the distribution of the number of active users within the network, which networks can exchange among themselves. Another information that could be exchanged is the scheduling policy or the provided data rate if it is constant. Then, knowing this information and the resource allocation policy, the transmitter network can predict, within some bounds, the values of the data rates at the receiver (network) end. Based on that, the transmitter BS can decide on the amount of resources to allocate to the transmitter. In this way, there would be throughput guarantees provided jointly with efficient utilization of network resources, where the latter then can be used to serve more users. This approach can be useful for the different network operating entities to better use their resources. The main message of the paper is that looking only at the distribution of the number of active users, and allowing the data rate to be within a (usually narrow) band can stabilize the end-to-end throughput and would lead to significantly lower resource wastage. Specifically, our main contributions can be summarized as:

- We propose an approach in which by only knowing the distribution of the number of active users on the receiving end of the multi-domain network one can determine the range in which data rates will lie, where the central value of the interval determines the end-to-end throughput.
- We propose an approach to determine the allocation policy when at the receiver the scheduling policy changes.
- Using realistic simulations, we show the advantages our approach brings in terms of both the throughput and efficient resource utilization against benchmarks.

II. Problem formulation

In this section, we present first the system model. This is followed by the problem setup.

A. System model

We consider a multi-domain network, consisting of multiple campus networks. This is illustrated in Fig. 1. In general, each campus network is operated by a different entity, and comprises the RAN and CN part, i.e., their operation resembles that of a traditional cellular network. In general, there can be multiple BSs associated with the unique CN in the same campus network. While the communication is between transmitters and receivers of the same campus network, there are no changes in the operation compared to traditional cellular networks. The challenge is faced when the receiver is within the coverage area of another campus network. In that case the CN of the transmitter-side network forwards through the Internet the data to the CN of the receiving-side network, which further forwards them to the corresponding BS by which the receiver is being served. This closes the one-way direction of the process (shown with the red curve in Fig. 1).

There are $|\mathcal{N}|$ campus networks. Within each campus network, we assume there are multiple BSs. The set of BSs within campus network i is $\mathcal{M}_i$. We consider mobile users, referred to as User Equipment (UE), within the coverage areas of each BS. The focus is both on the uplink and downlink. The set of active users within BS j belonging to campus network i is denoted by $\mathcal{L}_{i,j}$.

Each campus network has its own set of frequencies, and we assume that each BS operates on a fixed set of frequencies.[1] In all BSs, the Physical Resource Blocks (PRBs) are used as the unit of resource allocation on a per-slot basis [11]. Each PRB consists of 12 subcarriers. The slot duration is a function of the subcarrier spacing. Specifically, if the subcarrier spacing is 15 kHz (PRB width of 180 kHz), the slot duration is 1 ms. If the subcarrier spacing is 30 kHz (PRB width of 360 kHz), the corresponding slot duration is 0.5 ms. The slot duration decreases further ($2\times$) when switching to subcarrier spacing of 60 kHz, and another $2\times$ when switching to 120 kHz [11].[2] Different PRBs are assigned to different UEs within a slot. The assignment varies across slots. Consequently, scheduling needs to be performed across two dimensions, *frequency* and *time*. In total, there are K available PRBs in each BS of any campus network.[3]

[1]Nevertheless, having varying frequencies for a BS across time can be captured by our approach.

[2]As still there are no indications regarding actual structural changes in the resource allocation process in 6G, for that part in this work we use the notions from 5G.

[3]The analysis can be extended to different number of PRBs across different BSs/different campus networks.

UEs experience different channel conditions across different PRBs even within the same slot. This is captured by the parameter known as Channel Quality Indicator (CQI) [12], which attains values in the range $1-15$, with higher values for the better channel conditions. Because of the UE mobility and time-varying nature of the channels, per-PRB CQI (which is a function of Signal-to-Interference-Plus-Noise-Ratio (SINR)) changes from one slot to another, whose value depending on the Modulation and Coding Scheme (MCS) used sets the per-PRB rate [1]. To maintain analytical tractability, a simplifying assumption is made in this paper. Specifically, we assume that the BS splits the transmission power equally among all PRBs it transmits on, and that the channel characteristics for a UE remain static across all PRBs (identical CQI over all PRBs for a given UE), but change randomly (according to some distribution) from one slot to another, and are mutually independent among UEs (i.e., we are dealing with UEs with heterogeneous channel conditions). These assumptions reduce the resource allocation problem to the number of allocated PRBs and not to which PRBs are assigned to a UE.

The previous assumptions imply that in every slot, UE (i, j, k)[4], where $i \in \mathcal{N}$, $j \in \mathcal{M}_i$, $k \in \mathcal{L}_{i,j}$ will have a per-PRB rate $R_{i,j,k}$, i.e., the rate each assigned PRB brings to a UE. This per-PRB rate can be modeled as a discrete random variable with values in $\{r_1, r_2, \ldots, r_{15}\}$ (because of the 15 possible values of CQI), such that $r_1 < r_2 < \ldots < r_{15}$, with a Probability Mass Function (PMF) $p_{R_{i,j,k}}(x)$, which is a function of UE (i, j, k)'s CQI over time.

User activity: Users change their activity from *idle* to *active*. To capture this, we introduce the Bernoulli random variable with probability $q_{i,j,k}$ for UE (i, j, k). These values are independent across slots, users, BSs, and campus networks.

B. Problem setup

Given the lack of control and complete knowledge of the topology over the network of the receiver situated in a different campus network, the transmitter-network will not be able to infer the exact rate at which the data will reach the receiver. Therefore, as already mentioned in Section I, in this paper we consider *soft* throughput guarantees that should be provided to a communication session. This is formally defined as:

Definition 1. *The range of the data rate values in the uncontrollable network, i.e., the soft guarantee for the throughput, is* $[(1-\theta)U, (1+\theta)U]$*, where* U *is the central value of the range, whereas* θ *is the maximum deviation ratio.*

The value of the deviation ratio, θ, is usually small, and is controlled by the operator and determines the level of softness of the throughput guarantee.

There are two approaches in terms of the strictness of providing the soft throughput. In the first, the data rates should always fall within the interval $[(1-\theta)U, (1+\theta)U]$. In the second, for the vast majority of time the data rate should be within the aforementioned interval, and rarely the rate would be below the lower bound, $(1-\theta)U$. The ratio of time when the achieved data rate is not within the targeted interval is known as the *outage probability*, and is denoted by ϵ. In this paper, we follow the latter approach. The rationale behind this decision lies in the fact that it was already shown, in a different context [13], that relaxing the requirement of providing the rate from the targeted interval by only a small outage leads to considerably higher values of the lower and upper boundaries of the targeted interval. To summarize:

Definition 2. *The soft guarantee* $[(1-\theta)U, (1+\theta)]U$ *should be provided for* $(1-\epsilon)\cdot 100\%$ *of the time.*

In the next section, we provide the analysis of determining the central value of the targeted interval, U.

III. Analysis

The first step is to describe analytically the requirement for the soft throughput guarantee, and specifically its relation with the outage probability. In this section, we focus on a single UE as a transmitter in one campus network, and another UE as a receiver in another campus network, both sharing resources from their respective campus networks. Therefore, to simplify the notation, we replace the index (i, j, k) describing every user with the index t for the transmitting UE, and the receiving UE by the index r. Another assumption that we make in this section is that on the receiver side (the downlink) the resources by the BS of a campus network are allocated in a Round-robin fashion, although any other resource allocation policy can be captured by our approach. With the above changes and assumptions in mind, the soft throughput guarantee can be expressed as

$$\mathbb{P}\left((1-\theta)U \leq \frac{K_r R_r}{M_r} \leq (1+\theta)U\right) \geq 1-\epsilon, \qquad (1)$$

where $\frac{K_r}{M_r}$ denotes the amount of allocated resources (i.e., number of PRBs) to the user of interest, given the Round-robin resource allocation policy at the receiving network.[5] Note that there are two random variables in this constraint: R_r (per-PRB rate) and M_r (the number of active users in a slot at the receiver campus network), and the unknown to be determined is the maximum possible U that does not violate (1).

Given that data rates on the backhaul link (the link between the BS and the CN) and at the CN are considerably higher than in RAN [14], the overall throughput would be determined by the bottleneck link between the transmitting-node RAN and receiving-node RAN. Given that for most of the time the guaranteed rate in the receiving end is in the range $[(1-\theta)U, (1+\theta)U]$, the controllable data rate at the transmitting side should be tailored to that same range, i.e., the data rate at the transmitter part, $K_t R_t$, should be within

$$[(1-\theta)U, (1+\theta)U], \qquad (2)$$

[4]We denote every UE with the ordered pair (i, j, k), where i stands for the campus network, whereas j denotes the BS that provides service to user k.

[5]In fact, the number of assigned PRBs is an integer, and the more correct notation would be to use the floor function. Nevertheless, in order to simplify the notation, we omit the floor function in relation to the number of PRBs.

where K_t denotes the number of allocated PRBs to the transmitter by the corresponding BS of the transmitting-side campus network, whereas R_t, in line with simplifying the notation already introduced for the receiver side, is the per-PRB rate of the transmitter in the given slot. This implies that the number of allocated PRBs to the transmitter in a slot should be in the range

$$K_t \in \left[\frac{(1-\theta)U}{R_t}, \frac{(1+\theta)U}{R_t}\right], \tag{3}$$

Note that the highest loss due to the mismatch between the transmitter and receiver is $2\theta U$.

In order to fully describe the resource allocation policy on the uplink of the transmitting campus network side, i.e., (3), we need to determine the value of U. As a first step, (1) transforms into

$$\mathbb{P}\left(\frac{K_r R_r}{M_r} \leq (1+\theta)U\right) - \mathbb{P}\left(\frac{K_r R_r}{M_r} < (1-\theta)U\right) \geq 1-\epsilon. \tag{4}$$

Next, we proceed with deriving the two terms of the left-hand side (LHS) of (4). The first LHS term reduces to

$$\mathbb{P}\left(\frac{K_r R_r}{M_r} \leq (1+\theta)U\right) = \mathbb{P}\left(\frac{R_r}{M_r} \leq \frac{(1+\theta)U}{K_r}\right). \tag{5}$$

Note that there are two random variables in (5), R_r and M_r. Therefore, we need to condition upon one of the random variables. After conditioning upon the random variable M_r, (5) transforms into

$$\sum_{j=1}^{n_r} \mathbb{P}\left(\frac{R_r}{M_r} \leq \frac{(1+\theta)U}{K_r}\middle| M_r = j\right)\mathbb{P}(M_r = j), \tag{6}$$

where n_r denotes the scenario when all the users at the receiver-side BS are active, i.e., it is the number of all users currently residing within the coverage area of that BS. Denoting the PMF of the number of active UEs on the receiver campus network side by $p_{M_r}(j) = \mathbb{P}(M_r = j)$, after some calculus, we obtain

$$\sum_{j=1}^{n_r} \mathbb{P}\left(\frac{R_r}{j} \leq \frac{(1+\theta)U}{K_r}\right) p_{M_r}(j) = \sum_{j=1}^{n_r} \mathbb{P}\left(R_r \leq \frac{j(1+\theta)U}{K_r}\right) p_{M_r}(j). \tag{7}$$

Substituting (7) into (5), the following relation is obtained:

$$\mathbb{P}\left(\frac{K_r R_r}{M_r} \leq (1+\theta)U\right) = \sum_{j=1}^{n_r} F_{R_r}\left(\frac{j(1+\theta)U}{K_r}\right) p_{M_r}(j), \tag{8}$$

where F_{R_r} is the Cumulative Distribution Function (CDF) of the per-PRB rate of the receiver UE.

For the second LHS term of (4), we have

$$\mathbb{P}\left(\frac{K_r R_r}{M_r} < (1-\theta)U\right) = \mathbb{P}\left(\frac{K_r R_r}{M_r} \leq (1-\theta)U\right) - \mathbb{P}\left(\frac{K_r R_r}{M_r} = (1-\theta)U\right). \tag{9}$$

Further, we need to determine the two right-hand size (RHS) terms of (9). In that direction, the first RHS term, using a similar procedure as when deriving (8), yields

$$\mathbb{P}\left(\frac{K_r R_r}{M_r} \leq (1-\theta)U\right) = \sum_{j=1}^{n_r} F_{R_r}\left(\frac{j(1-\theta)U}{K_r}\right) p_{M_r}(j). \tag{10}$$

Following a similar procedure for the second RHS term of (9), with the distinction that it is a PMF and not a CDF, it follows that

$$\mathbb{P}\left(\frac{K_r R_r}{M_r} = (1-\theta)U\right) = \sum_{j=1}^{n_r} p_{R_r}\left(\frac{j(1-\theta)U}{K_r}\right) p_{M_r}(j). \tag{11}$$

The last missing piece of the puzzle is the PMF of the number of active users currently being active at the receiving-node BS, i.e., $p_{M_r}(j)$. Since M_r is the sum of Bernoulli distributions with different probabilities, $q_{i,j,k}$ (or its short version q_i), then M_r is subject to Poisson's Binomial distribution [15]. For this distribution, the probability of having j active users in a given slot is [15]

$$p_{M_r}(j) = \mathbb{P}(M_r = j) = \sum_{A \in F_j} \prod_{i \in A} q_i \prod_{k \in A^C} (1-q_k). \tag{12}$$

In (12), F_j is the set of all subsets of j users that can be selected from $\{1, \ldots, n_r\}$. As an example, if $n_r = 3$, for $j = 2$, $A_2 = \{\{1,2\},\{1,3\},\{2,3\}\}$. A^C is the complement of set A. F_j contains $\binom{n_r}{j}$ elements. Given the large number of possible combinations, it is difficult to compute (12) for large n_r. Nevertheless, there are ways to simplify the computation of the PMF $p_{M_r}(j)$ when none of the users is active at all times, which is a reasonable assumption. Due to space limitations, we refer the interested reader to [16] for more details on this.

Finally, substituting (12) into (8), (10), and (11), and the latter three equations into (4), we obtain:

Result 1. *The resource allocation policy at the transmitter (i.e., in the uplink) that provides soft throughput guarantees in a multi-domain network setup in a slot is given by (3), where U is obtained numerically as the largest value that satisfies the inequality*

$$\sum_{j=1}^{n_r} \left[F_{R_r}\left(\frac{j(1+\theta)U}{K_r}\right) - F_{R_r}\left(\frac{j(1-\theta)U}{K_r}\right) + p_{R_r}\left(\frac{j(1-\theta)U}{K_r}\right)\right] \sum_{A \in F_j} \prod_{i \in A} q_i \prod_{k \in A^C} (1-q_k) \geq 1-\epsilon. \tag{13}$$

While Result 1 is obtained numerically, the advantage of our approach is that inequality (13) needs to be computed only once (at the beginning) and then it is used across different slots. Nevertheless, this is a dynamic resource allocation policy due to the varying per-PRB rate (CQI changes in every slot in general), R_t, in (3).

The question that arises next is *how to determine the exact K_t from the range (3) in a slot?* Having the soft throughput guarantee and allowing the operator to choose the value from a range, instead of a fixed value, is an extra degree of freedom

our approach offers. The operator, depending on the channel conditions of the users, will compute the required resources for each UE, and if the request for resources is high, the operator will simply provide the rate $(1-\theta)U$ in that slot to the UE of interest. If resources are sufficient to satisfy the level of service for everyone, the operator can provide a higher rate (up to $(1+\theta)U$) to the UE of interest. In Section IV (Result 2), we propose an approach on how to determine what rate (and hence what K_t) from the feasible interval to provide in a slot depending on the average number of allowed PRBs for a UE.

IV. Consideration of Other Allocation Policies

In the analysis in the previous section, the resource allocation policy at the campus network of the receiver was Round-robin. In this section, we turn our attention to another resource allocation policy. Now, the assumption is that at the receiver campus network, the users are guaranteed a constant rate for almost 100% of the time. This rate is known as *consistent rate* [17].[6] Let us denote this data rate as $U_c(\epsilon)$.

There are two ways to determine $U_c(\epsilon)$ at the BS of the receiver campus network. In the first, the user can be guaranteed this rate in line with the Service Level Agreement with the operator. The other way would be the operator, based on the number of users within a given BS and their channel conditions, to derive the maximum value of this consistent rate for a given outage probability ϵ. For more details, the interested reader is referred to [17].

In line with the principal requirement in this work, that of soft performance guarantees for the throughput, the first allocation policy that we consider on the transmitter campus network is the one that provides the same consistent rate as in the receiver network, $U_c(\epsilon)$. If at the transmitter campus network, there are more favorable conditions in terms of the resources, stemming from fewer users and/or better channel conditions, providing a higher consistent rate than $U_c(\epsilon)$ is not a problem. In that case, the BS at the transmitter campus network will simply provide the necessary resources to maintain $U_c(\epsilon)$. This would be a *hard performance guarantee*. Here, we will focus on the more challenging scenario, the one when there are not always sufficient resources to provide $U_c(\epsilon)$ at the transmitter side.

Depending on the channel conditions of a user in a given slot, the required number of PRBs to provide $U_c(\epsilon)$ to that user is $\frac{U_c(\epsilon)}{r_l}$, where l is the CQI value of the user in a slot. As CQI can obtain one out of 15 possible values, the required number of PRBs to provide $U_c(\epsilon)$ in a slot is from the set

$$\left\{\frac{U_c(\epsilon)}{r_{15}}, \frac{U_c(\epsilon)}{r_{14}}, \ldots, \frac{U_c(\epsilon)}{r_1}\right\}.$$

In the setup with non-abundant resources, we assume that a user can get at most $\frac{U_c(\epsilon)}{r_1}$ PRBs in a slot (the requirement to experience a data rate of $U_c(\epsilon)$ with the worst possible channel conditions). The second assumption in this context is that on average (over time) the user should not "spend" more than K_m PRBs. This would correspond to a flexible resource allocation in which the operator would like to accommodate as many users as possible while satisfying their traffic requirements. If the consistent-rate policy is pursued in the transmitter campus network, a user with a PMF of per-PRB rates of $\{p_1, \ldots, p_{15}\}$ would need on average the following number of PRBs:

$$\mathbb{E}[K_c] = \sum_{l=1}^{15} \frac{U_c(\epsilon)}{r_l} p_l = U_c(\epsilon) \sum_{l=1}^{15} \frac{p_l}{r_l}. \tag{14}$$

As already mentioned, of interest and more challenging is the scenario in which $\mathbb{E}[K_c] > K_m$. Therefore, in Section V, we will compare the results of this consistent-rate approach against those of our approach (its adapted version for the transmitter campus network), when in the receiver campus network a consistent-rate policy is used, described next.

According to the policy proposed in this work, when the provided consistent rate $U_c(\epsilon)$ is provided at the receiver campus network, at the transmitted campus network the data rate should be in the range $[(1-\theta)U_c(\epsilon), (1+\theta)U_c(\epsilon)]$ for $1-\epsilon$ of the time for a given allowed deviation ratio of θ.[7] This leads to the required number of PRBs for a rate r_l, $l = 1, \ldots, 15$, to be in the range

$$\left[\frac{(1-\theta)U_c(\epsilon)}{r_l}, \frac{(1+\theta)U_c(\epsilon)}{r_l}\right]. \tag{15}$$

These assumptions entail the average required number of PRBs to be in the range

$$(1-\theta)U_c(\epsilon) \sum_{l=1}^{15} \frac{p_l}{r_l} \leq \mathbb{E}[K] \leq (1+\theta)U_c(\epsilon) \sum_{l=1}^{15} \frac{p_l}{r_l}. \tag{16}$$

To have a reliable comparison of performances of both approaches, we keep the same assumptions as previously, i.e., at most $\frac{U_c(\epsilon)}{r_1}$ PRBs can be assigned in a slot, and $\mathbb{E}[K] > K_m$. To satisfy the latter requirement, in line with the nature of our soft performance guarantee policy, we are allowed to reduce the number of assigned PRBs for a given CQI of the user to the lower bound of (15). The question that arises is *how to determine when to provide the lower bound of the data rate* $(1-\theta)U_c(\epsilon)$*?* The answer to this question can be obtained by observing the factor $\frac{p_l}{r_l}$ for a user. Specifically, this factor for a given CQI l determines the number of required PRBs, or equivalently, by how much the average number of allocated PRBs exceed K_m. Hence, for a given user in a slot, the lower bound of the data rate $(1-\theta)U_c(\epsilon)$ is provided for the CQI values that yield a high $\frac{p_l}{r_l}$. This policy can be summarized as:

Result 2. *Let $\frac{p_l}{r_l}$ be ranked in descending order. Let us denote the index of a term in this new array by j, implying that $j = 1$ corresponds to the CQI with the highest $\frac{p_l}{r_l}$. The rate*

[6] It has been shown in [17] that relaxing the requirement to guarantee a constant rate from 100% to 99% of the time increases the data rate considerably. Therefore, in this subsection, as far as this approach is concerned, we assume that a fixed data rate is provided with a probability of $1-\epsilon$.

[7] Note that in this case $U = U_c(\epsilon)$.

$(1-\theta)U_c(\epsilon)$ will be provided up to the CQIs (in the new ranking) satisfying

$$\max\left\{ j \left| (1-\theta)U_c(\epsilon)\sum_{k=1}^{j}\frac{p_k}{r_k} + U_c(\epsilon)\sum_{k=j+1}^{15}\frac{p_k}{r_k} \leq K_m \right. \right\}. \tag{17}$$

If after maximum j is determined from (17), the LHS is strictly smaller than K_m, the remaining resources are allocated to the user for other CQIs (starting from $j+1$) to provide a rate higher than $U_c(\epsilon)$ until the average resource utilization reaches K_m.

Essentially, what Result 2 says is that after ranking the CQIs according to the factor $\frac{p_l}{r_l}$ in descending order, the data rate $(1-\theta)U_c(\epsilon)$ is provided to the highest-ranked CQIs from this new set until there are enough resources (K_m) to guarantee $U_c(\epsilon)$ for the remaining (lower ranked) CQIs. If there are leftovers, then these lower-ranked CQIs will receive a rate higher than $U_c(\epsilon)$, up to $(1+\theta)U_c(\epsilon)$, until the average resource allocation reaches K_m.

In Section V, we will show that our approach considerably outperforms the consistent-rate approach in terms of efficient resource utilization while increasing the variability of the data rate insignificantly.

Other policies: Our approach is not confined only to policies presented previously, but it can be used jointly with other resource allocation policies in the receiver campus network as well. However, due to space limitations we omit those policies from further consideration in this work.

V. Performance Evaluation

In this section, first we describe the simulation setup. This is followed by outcomes from our approach and comparisons against benchmark models.

A. Simulation setup

In all the scenarios in this section, it is assumed that there are six users both on the transmitter- and receiver-campus network. There are two pairs of users that communicate in this multi-domain network; w.l.o.g. let us assume that these are user 1 at the transmission campus network and user 1 at the receiver campus network as the first communication pair, and user 2 at the transmission campus network with user 2 at the receiver campus network making the second communication pair. Their per-PRB rates and the corresponding PMFs are given in Table I, and are adapted from a public dataset [18] after some processing. Users are active at all times. As far as the other users are concerned, the results we present here are oblivious to their channel characteristics. Therefore, we omit showing their statistics.

The slot duration is 0.5 ms, implying a subcarrier spacing of 30 kHz. There are 12 subcarriers per PRB, leading to PRB widths of 360 kHz. The number of PRBs on both campus networks of interest is 273 [11]. The results are obtained from simulations run on MATLAB R2024b.

B. Round-robin at the receiver campus network

In this section, we show the very inefficient nature of Round-robin policy [11] in multi-domain networks[8]. To that end, we consider the communication between user 1 at the transmitter campus network and user 1 at the receiver campus network. Each one of them receives $1/6$ of the available PRBs at their campus networks. The evolution of their data rates across a span of 150 slots is shown in Fig. 2. As can be observed, the data rates of the transmitter are much higher than those of the receiver. This is a consequence of much better channel conditions of user 1 on the transmitter campus network than of user 1 on the receiver side, where the former experiences only channel conditions with very high values of CQI. On the other hand, user 1 on the receiver campus network experiences only channel conditions with medium CQI values. Fig. 2 also depicts the expectations of the data rates of these two users. This discrepancy in data rates leads to a total throughput of only 21 Mbps (equal to the average data rate in the bottleneck link, in this case in the receiver campus network). This means that resources on the transmitter campus network are not used efficiently.

Fig. 3 shows the evolution of data rates and their expectations for the other communication pair considered in this section, that of user 2 on the transmission campus network and user 2 on the receiver campus network. As user 2 on the transmission campus network experiences much better channel conditions than its counterpart in the receiver campus network, there is a considerable mismatch in the data rates between this communication pair as well. The throughput in this case is 24 Mbps, implying again that resource allocation at the transmitter campus network is not used efficiently.

Next, we consider the performance when our approach, presented in Section III, is used at the transmitter campus network. On the receiver side, Round-robin is used. The input parameters remain unchanged compared to the previous scenarios. Fig. 4 depicts the evolution of the data rate of user 1 at the receiver campus network (red curve) and U (obtained from (13)), where the latter pertains to user 1 at the transmitter campus network. The value of the deviation ratio is $\theta = 0.2$ and $\epsilon = 0$. Around the value of U (and the same bounds as for red curves) are the achievable rates for the transmitter over time. We are not showing them in order not to make the figure overcrowded. The data rates on both sides are much more closely matched now, leading to the same throughput as in Fig. 2, but with much more efficient utilization of network resources. Those "saved" resources at the transmitter campus network can be used to admit more users in the network or to improve the quality of service for the remaining users at the transmitter campus network.

Similar to Fig. 4, Fig. 5 portrays the data rate over a time span of 150 slots for user pair 2. Again, our approach is

[8]Note that both benchmarks used for performance comparison in this paper are in the original references [11] and [17] proposed in the context of single-domain public cellular networks. We adapt them in this work, as this is the first paper, to our best knowledge, that considers the throughput guarantees in multi-domain networks.

TABLE I

PER-PRB RATES AND THE CORRESPONDING PROBABILITIES FOR USERS 1 AND 2 ON THE TRANSMITTER CAMPUS NETWORK AND USERS 1 AND 2 ON THE RECEIVER CAMPUS NETWORK.

R (kbps)	48	73.6	121.8	192.2	282	378	474.2	612	772.2	874.8	1063.8	1249.6	1448.4	1640.6	1778.4
$p_{1,l}$ (tr.)	0	0	0	0	0	0	0	0	0	0	0	0	0.4	0.4	0.2
$p_{2,l}$ (tr.)	0	0	0	0	0	0	0	0	0	0	0	0	0.3	0.3	0.4
$p_{1,l}$ (rec.)	0	0	0	0	0	0.4	0.4	0.2	0	0	0	0	0	0	0
$p_{2,l}$ (rec.)	0	0	0	0	0.01	0	0.49	0.49	0.01	0	0	0	0	0	0

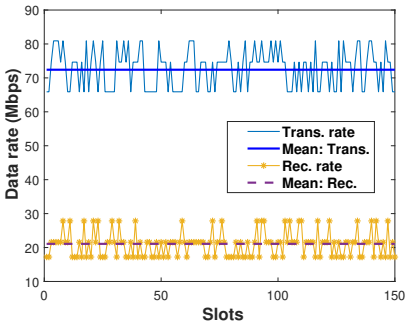

Fig. 2. The data rates at the transmitter (tr. 1) and receiver sides (rec. 1) and their means with Round-robin on both sides.

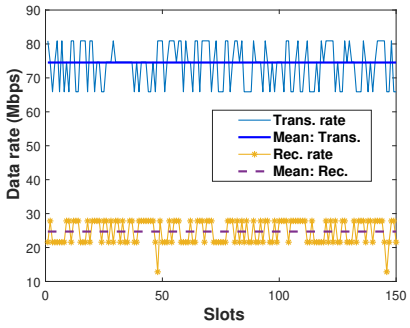

Fig. 3. The data rates at the transmitter (tr. 2) and receiver sides (rec. 2) and their means with Round-robin on both sides.

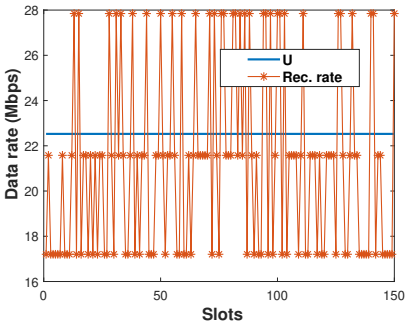

Fig. 4. The data rates at the transmitter (tr. 1) and receiver (rec. 1), with resource allocation at the transmitter according to our approach and Round-robin at the receiver.

shown to yield the same throughput (24 Mbps) as with Round-robin at both campus networks (see Fig. 3), but with saving considerable amount of resources at the transmitter campus network. The transmitter rates are bounded around the U value (flat blue line), but are not shown here for the same reasons as for Fig. 4. In this scenario, $\epsilon = 0.02$ (hence the spike in Fig. 5). Note that the same conclusion propagates across further slots, but due to better visibility we show results only for 150 slots.

C. Consistent rate at the transmitter campus network

Next, we proceed with the consistent-rate policy [17] at the transmitter campus network, while still having Round-robin as the resource allocation policy at the receiver campus network. The consistent rate is the maximum achievable data rate for the transmitter campus network for $\epsilon = 0.01$. Fig. 6 illustrates the resource utilization, expressed as a percentage, of user 1 at the transmitter side and user 1 at the receiver side. As can be observed, the resource utilization of the transmitter user 1 is much lower than receiver user 1 (for the latter the utilization is $\frac{100}{6}\%$ because there are six users on the receiver campus network) due to the much better channel conditions of transmitter user 1. Fig. 7 shows the results for communication pair 2. Similar conclusions hold as for communication pair 1 from Fig. 6.

From the previous two results, the natural question that arises is *whether providing the strict performance guarantee on the transmitter side is good enough in terms of efficient utilization of network resources?* It turns out that it is not. The approach relying on soft performance guarantees proposed in this paper provides better results, as will be shown next.

D. Consistent rate at the receiver campus network

Next, we compare the adapted version of our approach against the consistent-rate approach at the transmitter campus network, when at the receiver campus network the consistent-rate policy is being used. To that end, we keep the same two communication pairs as before. The consistent rate to be provided at the receiver campus network is 30 Mbps, with $\epsilon = 0.01$. On the transmitter side, we determine the data rates following Result 2. Table II shows the results for the coefficient of variation of data rates at the transmitter side with our policy for user 1 and user 2.[9] The important takeaway message from Table II is that for both users of interest on the transmitter campus network the variability of the data rates is very low, which is of high importance for users with stable throughput requirements.

In Table II, the coefficients of variation of data rates with the other two resource allocation approaches at the transmitter side, consistent rate and Round-robin, are also shown. The consistent-rate policy provides even lower values of the coefficient of variation of the data rates. This is expected as it is inherent to this policy to provide a constant rate at almost all times. Nevertheless, in the next scenario we show why allowing a slightly higher variation in data rates pays off in terms of resource allocation (by using our policy). The Round-robin resource allocation yields the worst results in terms of the variability, and the coefficient of variation with this policy is simply the variation of per-PRB rate of a user. The conclusions remain unchanged when varying the number of available PRBs, and similar trends are observed for other values of outage probability.

Having looked at the coefficient of variation of data rates with different resource allocation policies at the transmitter campus network, we proceed with investigating the excess of violation of the average PRB cap, K_m, when using our policy

[9]The coefficient of variation of a random variable is defined as the ratio of the standard deviation and the expectation of that random variable.

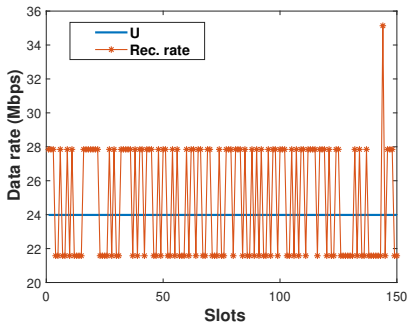

Fig. 5. The data rates at the transmitter (tr. 2) and receiver (rec. 2), with resource allocation at the transmitter according to our approach and Round-robin at the receiver.

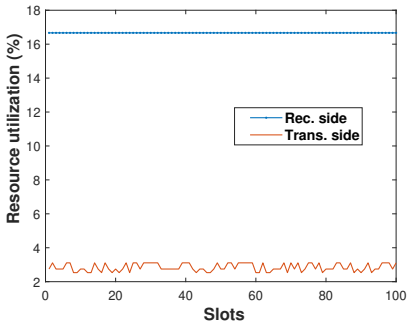

Fig. 6. Resource utilization of user 1 at the transmitter (consistent rate) and of user 1 at the receiver (Round-robin).

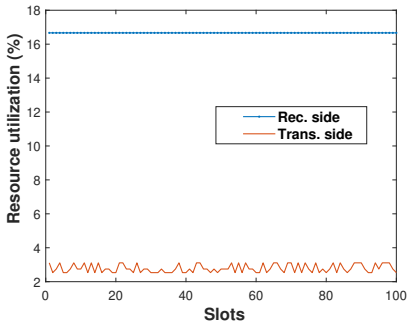

Fig. 7. Resource utilization of user 2 at the transmitter (consistent rate) and of user 2 at the receiver (Round- robin).

TABLE II
THE COEFFICIENT OF VARIATION FOR USERS 1 AND 2 AT THE TRANSMISSION CAMPUS NETWORK WITH THREE POLICIES

Policy	Our approach	Consistent rate	Round-robin
User 1	0.05	0.03	0.12
User 2	0.05	0.03	0.18

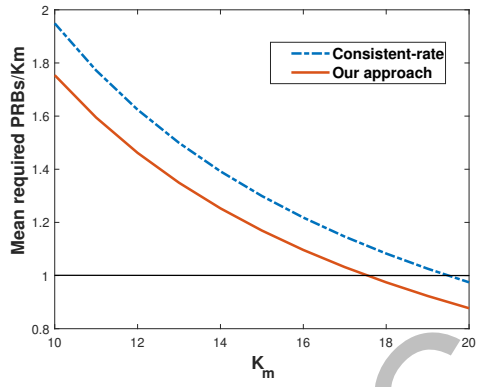

Fig. 8. The excess ratio of the mean number of PRBs required for user 1 and K_m at the transmission campus network as a function of the average PRB cap K_m for our policy and consistent-rate policy. The deviation ratio is $\theta = 0.1$, $\epsilon = 0.01$, and the data rate $U_c = 30$ Mbps.

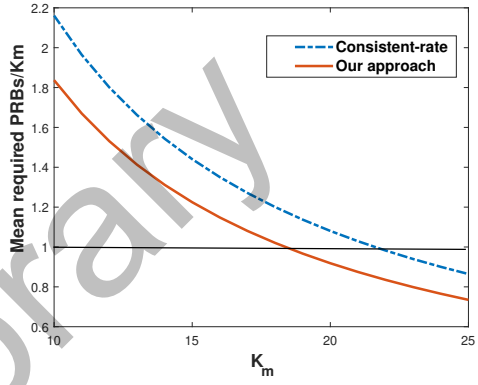

Fig. 9. The excess ratio of the mean number of PRBs required for user 2 and K_m at the transmission campus network as a function of the average PRB cap K_m for our policy and consistent-rate policy. The deviation ratio is $\theta = 0.15$, $\epsilon = 0.01$, and the data rate $U_c = 30$ Mbps.

(Result 2), and comparing it with the results obtained from the consistent-rate policy. As far as our policy is concerned, the allowed deviation ratio is $\theta = 0.1$ and $\epsilon = 0.01$, with a central value of 30 Mbps. Fig. 8 portrays the ratio of the average number of allocated PRBs and K_m, for different values of the latter for user 1 on the transmission campus network. This is a decreasing function in K_m, as expected, because the more the threshold is increased the more difficult to exceed it. In Fig. 8, the curve corresponding to the consistent-rate policy for user 1 is also shown. The consistent-rate value is $U_c = 30$ Mbps for $\epsilon = 0.01$. As can be observed, the excess ratio of the threshold K_m with the consistent-rate policy is higher than with our policy, up to 12%. This corroborates a more efficient resource utilization in a multi-domain network with our approach than with the consistent-rate policy. Similar results follow for user 2 on the transmission campus network as well.

Finally, we look at the excess ratio of the threshold K_m for user 2 on the transmission campus network when the deviation ratio is increased to $\theta = 0.15$. All the other parameters remain unchanged compared to the previous scenario. Fig. 9 depicts the results. Similarly to the previous scenario, our approach considerably outperforms the consistent-rate policy by reducing the excess ratio by 18%. It is again a decreasing function in K_m. Another interesting observation is that increasing the deviation ratio, θ, the gap between the two policies increases further due to the less restrictive requirement on the data rate at the transmitter side for our policy.

VI. Related Work

The problem of providing performance guarantees in cellular networks across specific metrics of interest has been known for a long time. In [19], the goal is to not exceed the maximum latency for almost all of the packets of a user. Performance guarantees in terms of reliability are considered in [20].

More related in spirit to this work, [17] and [21] focus on providing *hard* throughput guarantees to users belonging to the same and different use cases. For example, [17] proposes consistent rates for vast majority of the time. Another outcome from [17] is that relaxing the requirement on the time to guarantee the data rate from 100% of the time to a slightly lower value leads to significant improvements in the achievable data rates. However, this approach leads to an abundance of unused

resources, which makes the operation of cellular operators inefficient. While in [17] the analysis determines the maximum achievable rate for all users (the same data rate to everyone), in [21] the maximum achievable data rate is determined for each user separately, depending on their channel conditions. Furthermore, two approaches were considered in [21]. In the first, resources are reserved for each user from the very beginning, whereas in the second, the RAN resources are allocated on the fly. The latter approach was less costly for the cellular operators compared to [17]. However, despite the throughput guarantees of the aforementioned works, they hold only for single-domain networks, e.g., for public cellular networks, where the operator has full knowledge of the topology in the entire network across time. In the multi-domain network, this is not the case and consequently, the results from these related works cannot be applied.

To our best knowledge, there are no other works that tackle the problem of providing throughput guarantees in a multi-domain network setup. The closest work in spirit to ours is [13], where depending on the channel statistics of all the users, a range of values is determined with full resource utilization. So, similarly to the current work, there is no fixed data rate guaranteed, but it fluctuates across time between a range of values, and depending on the width of the feasible interval, the corresponding central data rate is determined. However, the approach in [13] is valid only over single-domain cellular networks where the same entity controls the overall network. In contrast, in the current work we consider the problem of providing throughput guarantees in a multi-domain network where the transmitting-side network obtains only partial information (on the distribution of the number of active users in the receiving-side network) from other campus networks. We perform the analysis that leads to the interval of data rates, which can then be used to determine the resource allocation policy on the transmitting side of the network.

VII. Conclusion

In this paper, we considered the problem of providing soft throughput guarantees in multi-domain networks, where the latter are managed by different operators. As different operators have no control over other domains, we propose the approach in which knowing the distribution of the number of active users at the receiver BS or their data rate, if the latter is constant for most of the time, the transmitter network can determine the range of values of the data rate it needs to provide, and based on that can perform the resource allocation accordingly. We showed that this improves the performance in terms of more efficient utilization of network resources than insisting on traditional strict constant rates or Round-robin, while preserving the end-to-end throughput.

As part of our future work, we plan to consider the problem of providing end-to-end guarantees in terms of latency across multi-domain networks.

Acknowledgment

This work was supported by the Federal Ministry of Education and Research of Germany (BMBF) under the project "6G-Life" with project identification number 16KISK002.

References

[1] F. Mehmeti and T. La Porta, "Reducing the cost of consistency: Performance improvements in next generation cellular networks with optimal resource reallocation," *IEEE Tran. on Mobile Computing*, vol. 21, no. 7, 2022.

[2] F. Mehmeti, V. T. Haider, and W. Kellerer, "Admission control for URLLC traffic with computation requirements in 5G and beyond," in *Proc. of IEEE/IFIP NOMS*, 2023.

[3] F. Mehmeti and T. La Porta, "Admission control for mMTC traffic in 5G networks," in *Proc. of ACM Q2SWinet 2021*.

[4] N. U. Ginige, K. B. Shashika Manosha, N. Rajatheva, and M. Latva-aho, "Admission control in 5G networks for the coexistence of eMBB-URLLC users," in *Proc. of IEEE VTC-Spring*, 2020.

[5] N. Fernández-Berrueta, S. Figueroa-Lorenzo, P. Bustamante, S. Arrizabalaga, and I. Velez, "5G performance measurements in mobility for the bus transportation system in an urban environment," *IEEE Access*, vol. 11, 2023.

[6] J. Wang, A. Jin, D. Shi, L. Wang, H. Shen, D. Wu, L. Hu, L. Gu, L. Lu, Y. Chen, J. Wang, Y. Saito, A. Benjebbour, and Y. Kishiyama, "Spectral efficiency improvement with 5G technologies: Results from field tests," *IEEE Journal on Sel. Areas in Communications*, vol. 35, no. 8, 2017.

[7] Z. Sun and Y. Jing, "Holographic MIMO NOMA communications: A power saving design," *IEEE Transactions on Wireless Communications*, vol. 23, no. 12, 2024.

[8] M. Hoffmann, G. Kunzmann, T. Dudda, R. Irmer, A. Jukan, G. Macher, A. Ahmad, F. R. Beenen, A. Bröring, F. Fellhauer, G. P. Fettweis, F. H. P. Fitzek, N. Franchi, F. Gast, B. Haberland, S. Hoppe, S. Joodaki, N. P. Kuruvatti, C. Li, M. Lopez, F. Mehmeti, T. Meyerhoff, L. Miretti, G. T. Nguyen, M. Parvini, R. Pries, R. F. Schaefer, P. Schneider, D. A. Schupke, S. Strassner, H. Stubbe, and A. M. Voicu, "A secure and resilient 6G architecture vision of the German flagship project 6G-ANNA," *IEEE Access*, vol. 11, 2023.

[9] M.-I. Corici, V. Gowtham, T. Magedanz, A. Prakash, and F. Schreiner, "NEMI: A 6G-ready AI-enabled autonomic network management system for open campus networks," in *Proc. of IEEE Globecom Workshops (GC Wkshps)*, 2022.

[10] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, "Toward 6G networks: Use cases and technologies," *IEEE Communications Magazine*, vol. 58, no. 3, 2020.

[11] ETSI, "5G NR overall description: 3GPP TS 38.300 version 15.3.1 release 15." www.etsi.org, 2018. Technical specification.

[12] F. Mehmeti and T. La Porta, "Analyzing a 5G Dataset and Modeling Metrics of Interest," in *Proc. of IEEE MSN*, 2021.

[13] F. Mehmeti, T. F. L. Porta, and W. Kellerer, "Efficient resource allocation with provisioning constrained rate variability in cellular networks," *IEEE Transactions on Mobile Computing*, vol. 23, no. 5, 2024.

[14] E. Goshi, F. Mehmeti, T. L. Porta, and W. Kellerer, "Modeling and analysis of mMTC traffic in 5G core networks," *IEEE Transactions on Network and Service Management*, 2024.

[15] Y. H. Wang, "On the number of successes in independent trials," *Statistica Sinica*, no. 3, 1993.

[16] B. K. Shah, "On the distribution of the sum of independent integer valued random variables," *Amer. Statistician*, vol. 27, no. 3, 1994.

[17] F. Mehmeti and C. Rosenberg, "How expensive is consistency? Performance analysis of consistent rate provisioning to mobile users in cellular networks," *IEEE Tran. on Mobile Computing*, vol. 18, no. 5, 2019.

[18] D. Raca, D. Leahy, C. J. Sreenan, and J. J. Quinlan, "Beyond throughput, the next generation: A 5G dataset with channel and context metrics," in *Proc. of ACM MMSys*, 2020.

[19] L. Dong and R. Li, "Latency guarantee service slice in 5G and beyond," in *Proc. of IEEE CCNC*, 2022.

[20] F. Mehmeti and T. La Porta, "Admission control for URLLC users in 5G networks," in *Proc. of ACM MSWiM 2021*.

[21] F. Mehmeti, A. Papa, W. Kellerer, and T. F. La Porta, "Minimizing rate variability with effective resource utilization in cellular networks," *IEEE Transactions on Mobile Computing*, vol. 23, no. 12, 2024.