FACULTY OF INFORMATION TECHNOLOGY AND ELECTRICAL ENGINEERING
DEGREE PROGRAMME IN WIRELESS COMMUNICATIONS ENGINEERING

MASTER'S THESIS

ADMISSION CONTROL IN 5G NETWORKS FOR THE COEXISTENCE OF EMBB-URLLC USERS

Author	Nipuni Uthpala Ginige
Supervisor	Prof. Nandana Rajatheva
Second Examiner	Dr. K. B. Shashika Manosha

October 2019

Ginige N. (2019) Admission Control in 5G Networks for the Coexistence of eMBB-URLLC users. University of Oulu, Faculty of Information Technology and Electrical Engineering, Degree Programme in Wireless Communications Engineering. Master's Thesis, 42 p.

ABSTRACT

In this thesis, we have considered the problem of admission control in 5G networks where enhanced mobile broadband (eMBB) users and ultra-reliable low-latency communication (URLLC) users coexist. Our aim is to maximize the number of admitted eMBB users to the system with a guaranteed data rate, while allocating power, bandwidth and beamforming directions to all URLLC users whose latency and reliability requirements are always guaranteed.

We have considered the downlink of a multiple-input single-output (MISO) network. We have considered orthogonal spectrum sharing between these two types of users. The maximum achievable data rate by an eMBB user is modelled using the Shannon equation. As the packet length of an URLLC user is small, to model the data rate of an URLLC user, we have used the approximation of Shannon's rate in short blocklength regime. Then, to further simplify and to obtain a lower bound for the short blocklength capacity equation, we have used the notion of effective bandwidth. This admission control problem is formulated as an ℓ_0 minimization problem. It is an NP-hard problem. We have used sequential convex programming to find a suboptimal solution to the problem.

Numerically we have shown the convergence of the algorithm. With numerical results, we have shown that number of admitted users increases with the increase of the total bandwidth of the system and maximum power of the base station. Further, it decreases with the increase of the target rate for eMBB users. Moreover, we have proven with the help of numerical results that the number of admitted users is decreasing with the increase of number of URLLC users in the system.

Keywords: 5G NR, eMBB and URLLC users, MISO, finite blocklength regime, effective bandwidth, bandwidth allocation, power allocation, sequential convex programming.

TABLE OF CONTENTS

FOREWORD

This thesis contains the results of the research carried to derive a solution for the problem of admission control in 5G downlink MISO network in which URLLC users and eMBB users coexist. This thesis has done as a part of the High5 and MOSSAF projects at the Center for Wireless Communications (CWC) of the University of Oulu, Finland and for the completion of Master's degree program in wireless communication engineering, University of Oulu, Finland.

First of all, I would like to take this opportunity to express my sincere gratitude to my supervisor Prof. Nandana Rajatheva, for his great support, guidance, and encouragement, given throughout my Master's studies. I am also thankful to Prof. Matti Latva-aho for selecting me to join with the University of Oulu and for providing me this opportunity to take part in the High5 and MOSSAF projects. Then, I wish to express my gratitude to my technical advisor Dr. Shashika Manosha for the guidance, support and encouragement he has provided me to complete my research. Also I would like to thank the project manager of High5 project, Dr. Pekka Pirinen, for the support given me to complete my research successfully. Moreover, I would like to thank all the Professors and Lecturers in University of Oulu, Finland who help me to complete the Master's degree successfully. I am also thankful to all the Professors and Lecturers in University of Moratuwa, Sri Lanka who help me to complete my Bachelor's degree successfully. Furthermore, I would like to thank all the teachers at Newstead Girls' College, Sri Lanka who help me to succeed in my primary and secondary education.

I want to express my special gratitude to my loving parents, for their love, kindness, support and encouragement provided throughout my life. Then I would like to thank my brother and sister for their love and kindness during my life. Finally, I would like to express my gratitude to my loving husband, for the love, encouragement and support given by him. Without him doing a Master's degree in a country which is far away from my mother country, would be a dream for me.

Oulu, 11th October, 2019

Nipuni Uthpala Ginige

LIST OF ABBREVIATIONS AND SYMBOLS

3GPP	3rd Generation Partnership Project
4G	Fourth Generation
5G	Fifth Generation
AWGN	Additive White Gaussian Noise
C-RAN	Cloud - Radio Access Network
CSI	Channel State Information
eMBB	Enhanced Mobile Broadband
FDD	Frequency Division Duplex
gNB	Next generation NodeB
HARQ	Hybrid Automatic Repeat Request
IoT	Internet of Things
JACoB	Joint Admission Control and Beamforming
LTE	Long Term Evolution
MISO	Multiple Input-Single Output
mMTC	Massive Machine-Type Communications
NOMA	Non-Orthogonal Multiple Access
NP - hard	Non-deterministic Polynomial-time hard
NR	New Radio
NSBPS	Null-Space-Based Preemptive Scheduler
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
PER	Packet Error Rate
PRB	Physical Resource Block
QoS	Quality of Service
SINR	Signal-to-Interference-plus-Noise Ratio
SNR	Signal-to-Noise Ratio
URLLC	Ultra-Reliable Low-Latency Communication

$a_i(.)$	Number of packets arrive to the base station form node i
b^e	Bandwidth allocated to all eMBB users
b_j^u	Bandwidth allocated to the jth URLLC user
C	Channel capacity
c_k	Small scale fading
D_k	Delay of kth user
D_{max}	Maximum packet delay or latency
D_q	Maximum queueing delay
D_{bsp}	Base station processing delay
D_{fa}	Frame alignment delay
D_{tx}	Transmission delay
D_{uep}	user processing delay
d_k	Normalized data symbol
E^B	Effective bandwidth
F	Total number of frames
f	Frame number
h_k	Channel vector of kth user

$\mathbf{h}_k^e$	Channel vector from base station to eMBB user k
$\mathbf{h}_j^u$	Channel vector from base station to URLLC user j
J	Total number of URLLC users
k	User index
K	Total number of eMBB users
M	Large positive constant
m	Number of messages
$\mathbf{m}_k$	Beamforming vector of kth user
N_0	Single-sided noise spectral density
n	Length of code word
p	Iteration index
p_k	Power of kth user
P_{total}	Maximum available power in the base station
R	Achievable rate
R_k^e	Maximum achievable rate of kth eMBB user
R_j^u	Maximum achievable rate of jth URLLC user
R_{target}	Target rate for eMBB users
r_k	Distance from base station to kth user
r_0	Far field reference distance
S	Set of near by nodes
s_k	Auxiliary variable of kth user
T	Number of transmit antennas of the base station
T_f	Duration of one frame
$\mathcal{U}$	Set of users
$\mathcal{U}_o$	Set of admitted users
$\mathcal{U}_e$	Set of all eMBB users
$\mathcal{U}_u$	Set of all URLLC users
$\mathbf{u}_k$	Normalized beamformer of kth user
V	Channel dispersion
w_k	Additive white Gaussian noise
w_k^e	Additive white Gaussian noise at eMBB user k
w_j^u	Additive white Gaussian noise at URLLC user j
$\mathbf{X}$	Transmitted signal vector
y_k	Received signal vector of user k
α	Path loss exponent
β_k	Interference plus noise experienced by the kth eMBB user
γ_k^{target}	Minimum SINR requirement for kth user
$\gamma_k^{e,th}$	SINR threshold of kth eMBB user
γ_k^e	SINR of kth eMBB user
γ_j^u	SINR of jth URLLC user
$\gamma_j^{u,th}$	SINR threshold of jth URLLC user
δ	Small positive constant
ϵ	Overall reliability requirement
ϵ_c	Transmission-error probability
ϵ_q	Queuing-delay violation probability
ϵ_k^q	Queueing delay violation probability of user k

η_k	Probability of non-empty buffer
θ_k	QoS exponent for kth user
λ	Packet arrival rate
λ_k	Arrival rate for kth user
μ	Number of bits contained in each packet
ρ_i	Predefined threshold
σ_e^2	Noise variance of eMBB users
$\sigma_{j,u}^2$	Noise variance of jth URLLC user
τ	Duration for data transmission in one frame
τ^k	Convex trust region

$\sqrt{X}$	Square root of X
$\mathbb{C}^n$	Set of complex n-vectors
$\mathbb{E}(.)$	Expectation
$\exp(.)$	Exponential function
$\mathbb{R}^n$	Set of real n-vectors
ln	Natural logarithm
$\max(.)$	Maximum
$\min(.)$	Minimum
$\sum$	Sum operator
Q^{-1}	Inverse of Q function
$\nabla f(x)$	Gradient of function f at x
X^H	Conjugate transpose (Hermtian) of matrix X
X^T	Transpose of matrix X
$\mathrm{Im}(z)$	Imaginary part of a complex number z
$\lvert X\rvert$	Absolute of vector X
$\parallel x \parallel_0$	Cardinality of vector x (number of nonzero elements)
$\parallel x \parallel_2$	ℓ_2 norm of vector x
$\forall$	For all
log	Logarithm
$\mathbf{CN}(m, K)$	Circularly symmetric complex Gaussian distribution with mean m and variance $k/2$ per dimension

1 INTRODUCTION

Fifth generation (5G) new radio (NR) is the latest radio access technology which introduced after fourth generation (4G) long term evolution (LTE) by 3rd generation partnership project (3GPP). The vision of 5G is the possibility of accessing information and sharing data anytime and anywhere to anyone and anything [1, 2]. 5G can provide high peak data rates, low latency, high reliability, higher user mobility, higher connection density, higher throughput and many more [1,2]. 5G NR plays a vital role in applications such as autonomous vehicle control, smart city, high-speed trains, virtual and augmented reality, emergency communication, factory automation, large outdoor events, media applications, remote surgery and examination and inside shopping malls [1].

5G NR supports three main use cases. They are enhanced mobile broadband (eMBB), ultra-reliable low-latency communications (URLLC) and massive machine-type communications (mMTC).

- Enhanced mobile broadband (eMBB) - Provides high data rate, high spectral efficiency, low latency as well as high coverage.
- Ultra-reliable and low-latency communication (URLLC) - Provides ultra reliability and low latency.
- Massive machine-type communication (mMTC) - Supports for higher traffic density and scalable connectivity when the number of devices is increasing [1].

There are practical scenarios where eMBB users and URLLC users share the same resource. As an example in a shopping mall or at a large outdoor event there are applications which need low latency and high reliability such as online financial transactions or security applications, and there are applications which needs high data rate such as enjoying a music video by a person. In a scenario like a drone race or self-driving vehicle race, low latency is essential for drone communication and autonomous vehicle control. In that scenario, in the audience there can be some people accessing the internet who need high data rates.

According to 3GPP, quality of service (QoS) requirements of URLLC is ultra-high reliability ($1-1\times10^{-5}$ success probability) and low transmission latency of 1 ms, whereas eMBB requires high data rates of 1 Gbps [3]. Coexisting eMBB and URLLC users in the same resource is a difficult task since simultaneously achieving high data rates for eMBB users and the ultra reliability and low latency for URLLC users becomes a challenging scheduling task since there is a trade-off between latency, reliability and achieving high data rates.

Admission control in wireless networks can be interpreted as finding the maximum amount of traffic or maximum number of users that can be admitted simultaneously to the system while efficiently using the available resources and satisfying QoS requirements. Another interpretation for admission control is deciding the newly arriving traffic or newly arriving user can be admitted to the system with the available resources of the system and the QoS requirements of entering traffic.

Effective bandwidth is the minimum amount of the bandwidth required to satisfy QoS requirements [4]. If the maximum achievable rate of the URLLC user is greater than or equal to the effective bandwidth, which was derived using the reliability and latency values, we can say that reliability and latency requirement of the URLLC user is guaranteed [5].

We have derived our problem for the downlink of a multi-user multiple-input single-output (MISO) network. We suggest orthogonal spectrum sharing between eMBB users and URLLC users to coexist them. The maximum achievable data rate by an eMBB user is modelled using the Shannon equation. As the packet length of an URLLC user is small, to model the data rate of an URLLC user we have used the approximation of Shannon's rate in short blocklength regime. Then, to further simplify and to obtain a lower bound for the short block length capacity equation, we have used the notion of effective bandwidth. We have formulated the problem as an ℓ_0 minimization problem. It is an non-deterministic polynomial-time hard (NP-hard) problem. Therefore, we need to make some approximations to find an optimal solution. We have used sequential convex programming to solve the problem.

1.1 Motivation

Coexistence of eMBB and URLLC applications is important since resources are limited. With the growth of wireless users and limitation of resources and also, the QoS requirements of users, it is important to know the maximum number of users that can be admitted to the system. Therefore, admission control plays a vital role in wireless networks.

There are many literature which speak about scheduling algorithms for the coexistence of eMBB and URLLC applications [3, 6–8]. Moreover, there are existing literature which intend to find a solution for admission control problem in wireless networks [9–12]. Most of the scheduling algorithms for the coexistence of eMBB and URLLC in literature suggest puncturing eMBB users, in order to give priority to URLLC users and satisfy their reliability and latency requirements. Therefore, it is essential to know the possible number of eMBB users that can be supported by the system while meeting the reliability and latency requirements of URLLC users.

However, no research has been found that speaks about admission control problem in the wireless network in which eMBB users and URLLC users coexist. Therefore, it motivates us to find a solution for the problem of admission control in 5G networks in which eMBB users and URLLC users coexist.

1.2 Objectives of the thesis

One purpose of this thesis is to study existing literature about scheduling algorithms for networks in which eMBB users and URLLC users coexist and existing literature about solutions for the admission control problem in wireless network. The main purpose of this thesis is to investigate a method to coexist eMBB and URLLC users and a solution for the admission control problem of eMBB users in 5G networks in which eMBB users and URLLC users coexist.

We have proposed orthogonal spectrum sharing between eMBB and URLLC users to coexist them. Furthermore, we have contributed an algorithm to solve the admission control problem of eMBB users in 5G networks in which eMBB users and URLLC users coexist. During the derivation of the algorithm, our ultimate goal is to maximize the number of admitted eMBB users to the system, who have sufficient data rate, while allocating power, bandwidth and beamforming directions to all URLLC users whose latency and reliability requirements are always guaranteed.

1.3 Structure of the thesis

This thesis consists of five chapters and the rest of the thesis is organized as follows:

- Chapter 2 - In this chapter, we provide the background knowledge related to the downlink multi-user MISO network, Shannon's rate and approximation for Shannon's rate in finite blocklength regime. This chapter includes about 5G NR and details of two of its use cases: eMBB and URLLC as well. Then in this chapter, we briefly describe about effective bandwidth concept for URLLC which we have used to derive the lower bound for signal-to-noise ratio (SNR) threshold for URLLC users. We show some of the examples from literature, where the idea of effective bandwidth has used for SNR and achievable rate calculations of URLLC users. Then we present brief description of admission control in the wireless networks and sequential convex programming. Finally, in this chapter we summarize existing literature on scheduling algorithms for the coexistence of eMBB and URLLC and existing literature which provides solutions for the admission control problem in the wireless networks.

- Chapter 3 - This chapter provides our solution for the problem of admission control for eMBB users in the coexistence of eMBB and URLLC in 5G network. In this chapter, we describe our system model and problem formulation. Furthermore, this chapter includes the derivation of the algorithm of our solution for the problem of admission control for eMBB users in 5G networks in which eMBB users and URLLC users coexist. At the end of this chapter, we introduce the algorithm of our solution.

- Chapter 4 - This chapter discusses about the simulations carry on to prove the correctness of our solution for the problem of admission control for eMBB users in 5G networks in which eMBB users and URLLC users coexist. It also provides the numerical results which we obtain to illustrates the performance and effectiveness of our solution.

- Chapter 5 - This chapter includes a conclusion for this thesis. Moreover, it provides potential future research directions.

2 BACKGROUND THEORY AND LITERATURE REVIEW

In this chapter, we present the background knowledge related to the downlink multi-user MISO network, Shannon's rate and approximation for Shannon's rate in finite blocklength regime. Furthermore, this chapter includes about 5G NR and details of two of its use cases: eMBB and URLLC. Then, we briefly describe about effective bandwidth concept for URLLC which we have used to derive the lower bound for SNR threshold of URLLC users. We give some of the examples from literature, where the idea of effective bandwidth has used for SNR and achievable rate calculations of URLLC users. Also, we present a brief description of admission control in the wireless networks and sequential convex programming. Finally, in this chapter we summarize existing literature on scheduling algorithms for the coexistence of eMBB and URLLC and existing literature which provides solutions for the admission control problem in the wireless networks.

2.1 Downlink multi-user MISO communication

Let us consider K number of users with single receive antenna and a base station with T transmit antennas. Figure 2.1 illustrates the downlink MISO network.

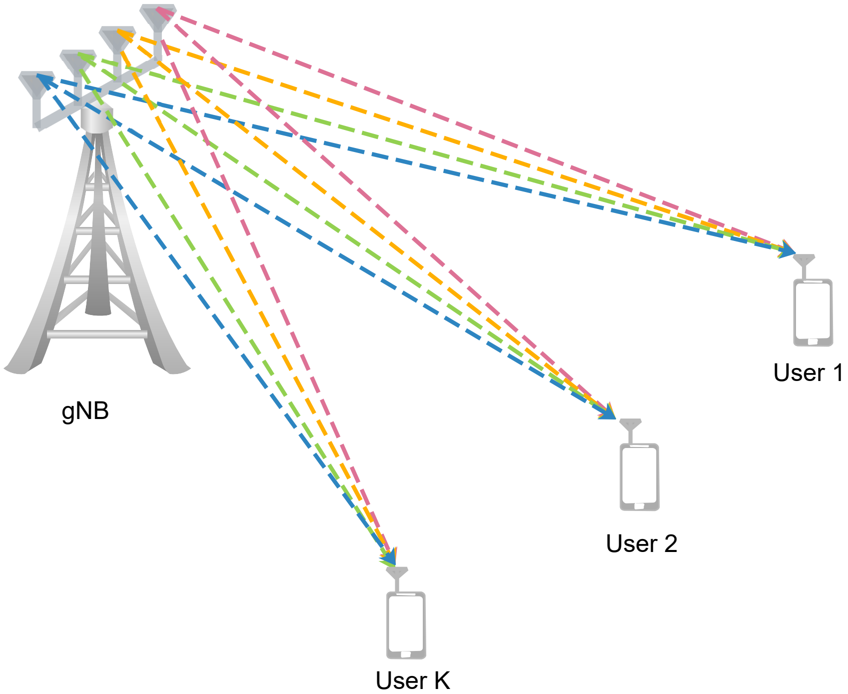

Figure 2.1. Illustration of the downlink MISO network.

The received signal vector at symbol time m can be expressed as

$$y_k[m] = \mathbf{h}_k^H \mathbf{X}[m] + w_k[m] = \mathbf{h}_k^H \mathbf{u}_k \sqrt{p_k} d_k[m] + \sum_{i=1, i\neq k}^{K} \mathbf{h}_k^H \mathbf{u}_i \sqrt{p_i} d_i[m] + w_k[m], \quad (1)$$

where $y_k \in \mathbb{C}$ is the received signal of kth user, $\mathbf{X} \in \mathbb{C}^T$ is the transmitted signal vector, $\mathbf{u}_k \in \mathbb{C}^T$ is the normalized beamformer, d_k is the normalized data symbol, p_k is the power of kth user, $\mathbf{h}_k \in \mathbb{C}^T$ is the channel vector of kth user and $w_k \sim \mathcal{CN}(0, N_0)$ is complex white Gaussian noise of kth user and N_0 is the single-sided noise spectral density. The beamforming vector $\mathbf{m}_k \in \mathbb{C}^T$ can be written as

$$\mathbf{m}_k = \sqrt{p_k}\mathbf{u}_k. \tag{2}$$

It is needed to minimize the transmit power while satisfying user specific signal-to-interference-plus-noise ratio (SINR) targets. In order to find optimum beamformer which is satisfying user-specific SINR targets, formulation of optimization problem is as follows [13]:

$$\begin{aligned} \text{minimize} \quad & \textstyle\sum_{k=1}^{K} \parallel \mathbf{m}_k \parallel_2^2 \\ \text{subject to} \quad & \frac{|\mathbf{h}_k^H \mathbf{m}_k|^2}{\sum_{i=1,i\neq k}^{K} |\mathbf{h}_k^H \mathbf{m}_i|^2 + N_0} \geq \gamma_k^{target}, \;\; k = 1, \ldots, K && \text{(3a)} \\ & \textstyle\sum_{k=1}^{K} \parallel \mathbf{m}_k \parallel_2^2 \leq P_{total}, && \text{(3b)} \end{aligned}$$

where γ_k^{target} is the minimum SINR requirement for kth user and P_{total} is the maximum available power in the base station and the optimization variables are $\mathbf{m}_k \in \mathbb{C}^T$ for $k = 1, \ldots, K$. We can prove that $\sum_{k=1}^{K} \parallel \mathbf{m}_k \parallel_2^2$ is equal to $\sum_{k=1}^{K} p_k$.

Equivalent second-order cone programming formulation for the optimization problem (3) is as follows [9]:

$$\begin{aligned} \text{minimize} \quad & \textstyle\sum_{k=1}^{K} \parallel \mathbf{m}_k \parallel_2^2 \\ \text{subject to} \quad & \mathbf{h}_k^H \mathbf{m}_k \geq \sqrt{\gamma_k^{target} \textstyle\sum_{i=1,i\neq k}^{K} |\mathbf{h}_k^H \mathbf{m}_i|^2 + \gamma_k^{target} N_0}, \;\; k = 1, \ldots, K && \text{(4a)} \\ & \textstyle\sum_{k=1}^{K} \parallel \mathbf{m}_k \parallel_2^2 \leq P_{total} && \text{(4b)} \\ & Im(\mathbf{h}_k^H \mathbf{m}_k) = 0, \forall k, && \text{(4c)} \end{aligned}$$

where the optimization variables are $\mathbf{m}_k \in \mathbb{C}^T$ for $k = 1, \ldots, K$.

2.2 Channel capacity

In a discrete memoryless channel when m number of messages have mapped to n length code word, maximum achievable rate can be obtained as

$$R = \frac{\log_2 m}{n}. \tag{5}$$

When the average probability of error is equal to zero, maximum achievable rate, in other words channel capacity can be obtained as

$$C = \max p_x I(\mathbf{X}; \mathbf{Y}), \tag{6}$$

where $\mathbf{X}$ is input vector, $\mathbf{Y}$ is the output vector, $I(X;Y)$ is the mutual information between $\mathbf{X}$ and $\mathbf{Y}$ and maximum has taken over all possible input distributions $p(x)$ [14]. Shannon has proven that the average probability of error tends to zero when the code length $n \rightarrow \infty$. For a Gaussian channel, channel capacity is given by [14]

$$C = \max p_x I(x; y) = \frac{1}{2}\log_2(1 + SINR). \tag{7}$$

For complex baseband additive white Gaussian noise (AWGN) channel, channel capacity becomes [14]

$$C = \log_2(1 + SINR) \quad bits/Hz. \tag{8}$$

2.2.1 Approximation for Shannon's rate in finite blocklength regime

Shannon's rate cannot apply when the code length is short. Therefore, the authors in [15] have introduced an approximation for maximum achievable rate in finite blocklength regime and it can be expressed as

$$C = \log_2(1 + SINR - \sqrt{\frac{V}{n}}Q^{-1}(\epsilon_c)), \tag{9}$$

where V is the channel dispersion, n is the blocklength, ϵ_c is the transmission error probability and Q^{-1} is the inverse of Q function.

2.3 5G NR

5G NR is the latest radio access technology which is introduced after 4G LTE by 3GPP. The vision of 5G is the possibility of accessing information and sharing data anytime and anywhere to anyone and anything [1, 2]. Aim of 5G is to fulfill many applications which need high peak data rates, low latency, high reliability, higher user mobility, higher connection density, higher throughput and many more [1,2]. There are eight requirements for 5G they can be listed as follows [16].

- Up to 1 Gbps data rate
- 1 ms latency
- 1000x bandwidth per unit area
- Up to 100x number of connected devices per unit area
- 99.999% reliability
- 100% coverage
- 90% energy efficiency
- Low power consumption

Let us consider some applications that 5G is playing an important role. One application is autonomous vehicle control. Vehicle-to-infrastructure communication, vehicle-to-vehicle communication, vehicle-to-people communication and vehicle-to-sensors communication happens in autonomous vehicle control. In autonomous vehicle control, low latency, high reliability and high availability are essential. Emergency communication is another application which 5G is playing an important role. In an emergency situation, high availability and high energy efficiency are essential things to consider. 5G is playing a significant role in factory automation. For factory automation, high reliability and low latency are important. Another application of 5G is high-speed train. In this context supporting higher user mobility is important. 5G is important in large outdoor events. In this scenario higher throughput and supporting higher traffic density is important. 5G is important for massive amount of geographically spread devices. For those devices higher energy efficiency is important. Media on demand is another application of 5G. In this case, higher data rates and low latency is important to give better performance for users. A very low end-to-end latency and ultra-reliable communications is needed for remote surgery and examination, which is another application of 5G. Main challenges in a shopping mall are providing secure connection for financial transactions, for surveillance cameras and also providing connection for large number of users. This is where 5G comes to play since those requirements need high reliability and high availability. Smart city is another application of 5G. A smart city connects people with the environment in a smart way to ease their day to day life. To provide smart services which are varying in wide range it needs to provide high data rates, low latency as well as high throughput. 5G plays an essential role inside a stadium since in a stadium it needs to support high traffic density. Teleprotection in smart grid network requires high reliability and low latency, and therefore 5G is important for those kinds of scenarios. Another important application of 5G is virtual and augmented reality. Virtual and augmented reality needs very high data rates and low latency [1].

5G NR supports three main use cases. They are enhanced mobile broadband (eMBB), ultra-reliable low-latency communications (URLLC) and massive machine-type communications (mMTC). In this thesis, we are talking about the coexistence of eMBB and URLLC users in the same resource.

2.3.1 eMBB

Enhanced mobile broadband is one of the use cases of 5G. eMBB is the evolved version of 4G which provides higher data rates and a better experience for users. eMBB delivers higher data rates, enhanced connectivity and higher user mobility. eMBB enables 360° video streaming, virtual reality and augmented reality media applications, etc. [17]. eMBB services maximize the data rate while providing moderate reliability and 10^{-3} packet error rate (PER) [18]. The maximum achievable rate can be calculated using Shannon's rate equation.

2.3.2 URLLC

Ultra-reliable low latency communication is another use case of 5G. URLLC applications guarantee 1 ms or lower latency and 1×10^{-5} or smaller reliability. URLLC supports applications like autonomous driving and factory automation [18].

URLLC requires sending short packets. The achievable rate of URLLC cannot be derived using Shannon's capacity since URLLC has short packets that transmit in small time duration. Therefore, the block length of channel codes of URLLC user is short [19]. When calculating the maximum achievable rate for URLLC applications, it needs to consider the approximation for Shannon's rate for finite block-length regime which mentions in the Equation (9).

2.4 Effective bandwidth for URLLC

Effective bandwidth is the minimum amount of bandwidth required to satisfy QoS requirements [4]. According to the authors in [20], effective bandwidth can be interpreted as the minimum required constant service rate to serve a random arrivals in order to satisfy queueing delay requirements and reliability requirements.

The authors in [5] have proved that the concept of effective bandwidth can be used in the finite block length regime for a Poisson arrival process to guarantee reliability and latency requirement of URLLC users. Effective bandwidth of kth user for stationary packet arrival process can be defined as [5]

$$E_k^B = \lim_{F\to\infty} \frac{1}{FT_f\theta_k} \ln\{\mathbb{E}[\exp(\theta_k \textstyle\sum_{f=1}^{F} \sum_{i\in S} a_i(f))]\} \quad packets/s, \tag{10}$$

where E_k^B is effective bandwidth for user k, f is the frame number, F is the total number of frames, θ_k is the QoS exponent for kth user, T_f is the duration of one frame, S is the set of near by nodes and $a_i(.)$ is the number of packets arrive to the base station form node i.

URLLC users usually have a delay less than 1 ms. Thus, their channel delay is less than the channel coherence time. Therefore channel is quasi-static and service rate is constant. When the constant service rate is equal to the effective bandwidth, queueing delay violation probability can be expressed as

$$Pr\{D_k(\infty) > D_q\} \approx \eta_k \exp\{-\theta_k E_k^B(\theta_k) D_q\}, \tag{11}$$

where η_k is the probability of non-empty buffer, D_k is the delay of kth user, and D_q is the maximum queueing delay.

To satisfy the queueing delay violation probability requirements, the upper bound of Equation (11) should be equal to the queueing delay violation probability requirement (ϵ_k^q) of user k. Therefore when $\eta_k \leq 1$, the queueing delay violation probability requirement is as follows:

$$\epsilon_k^q = \exp\{-\theta_k E_k^B(\theta_k) D_q\}. \tag{12}$$

Thus, the effective bandwidth which satisfies the queueing delay violation probability and queueing delay requirements can be expressed as

$$E_k^B = \frac{1}{D_q\theta_k} \ln \frac{1}{\epsilon_k^q}. \tag{13}$$

For Poisson arrival process with λ_k arrival rate for kth user, QoS component (θ_k) for the kth user is given by

$$\theta_k = \ln\left[\frac{T_f \ln \frac{1}{\epsilon_k^q}}{\lambda_k D_q} + 1\right]. \tag{14}$$

Finally, the authors in [5] have derived the effective bandwidth, which satisfies the queueing delay violation probability and queueing delay requirements is as follows:

$$E_k^B = \frac{\ln \frac{1}{\epsilon_k^q}}{D_q \ln [\frac{T_f \ln \frac{1}{\epsilon_k^q}}{\lambda_k D_q} + 1]} \quad packets/s. \tag{15}$$

In [19], an optimal resource allocation method for multi-user URLLC network, such that the latency and reliability requirements are guaranteed is considered. The concept of effective bandwidth is used to find the service rates for URLLC users, which can ensure the reliability and latency requirements of URLLC users. The authors in [19] suggest to drop packets proactively to acquire latency and reliability requirements, when the latency and reliability requirements cannot satisfy with the available resources.

Resource allocation problem for URLLC users is considered in [21]. The problem is formulated to maximize the energy efficiency by optimizing antenna configuration with the optimal bandwidth and power allocation. The effective bandwidth concept is used to find the service rates for URLLC users which satisfies the reliability and latency requirements.

The authors in [20] mentioned that the concept of effective bandwidth can use to ensure queuing delay requirements of URLLC users for resource allocation problems.

The performance of non-orthogonal multiple access (NOMA) for URLLC users is analysed in [22]. The concept of effective bandwidth is used to model the queueing delay and reliability for the performance analysis.

The resource allocation problem for URLLC users using unsupervised deep learning is considered in [23]. The authors in [23] jointly solved the optimization problem of bandwidth allocation and power allocation. In the optimization problem, the total bandwidth of the system is minimized while satisfying reliability and latency requirements of URLLC users. The concept of effective bandwidth is used to find a service rate which satisfy the reliability and latency requirements of URLLC users.

2.5 Coexistence of eMBB and URLLC in 5G

According to 3GPP, quality of service (QoS) requirements of URLLC is ultra-high reliability and low transmission latency, whereas eMBB requires high data rates [3]. There are practical situations which needs to share the same resources by eMBB users and URLLC users. Coexisting eMBB and URLLC users in the same resource is a difficult task since simultaneously achieving high data rates for eMBB users and the ultra-reliability and low latency for URLLC users becomes a challenging scheduling task since there is a trade-off between latency, reliability and achieving high data rates. There is a large number of published studies to introduce scheduling algorithms for the coexistence of URLLC and eMBB users in the 5G network to address this challenge.

A null-space-based preemptive scheduler (NSBPS) to coexist eMBB and URLLC users is introduced in [6]. An algorithm to satisfy the URLLC latency requirements while achieving the maximum rate for eMBB users has proposed. The optimization problem is formulated in a way such that maximize the rate of eMBB users and minimize the latency of URLLC users. The latency (D_{max}) is calculated as a combination of queuing delay

(D_q), base station processing delay (D_{bsp}), frame alignment delay (D_{fa}), transmission delay (D_{tx}) and user processing delay (D_{uep}) as follows:

$$D_{max} = D_q + D_{bsp} + D_{fa} + D_{tx} + D_{uep}. \tag{16}$$

URLLC packet arrival is modeled as a Poisson process, and the M/M/1 queueing model is used to calculate the latency in [6]. M/M/1 is Poisson input (M-Markovian) / exponential service (M) / single-server (1) queueing model. NSBPS scheduler works as follows. If there are not enough resources for the URLLC users, the proposing NSBPS scheduler immediately searches for an eMBB user who is currently using resources and whose direction is most align to a predefined reference subspace. Then the scheduler directs the selected eMBB user into the reference subspace. At the URLLC user side, it changes the orientation of its transmit beamforming vector into one possible null space of the reference subspace. This procedure is illustrated in Figure 2.2 [6].

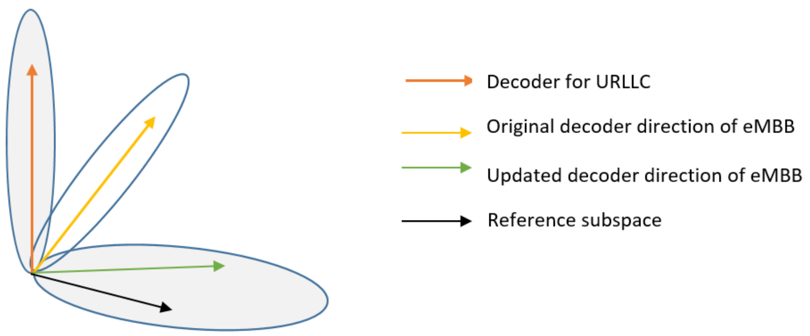

Figure 2.2. Illustration of NSBPS scheduler.

A low-complexity packet scheduling algorithm for URLLC and eMBB coexistence in 5G has proposed in [7]. The knowledge of traffic, latency, hybrid automatic repeat request (HARQ), radio channel and control channel is used to reduce the queuing delay and increase the reliability. In the resource allocation problem, the objective is to maximize the sum of the URLLC traffic and eMBB traffic while satisfying the reliability and latency requirements of URLLC users. Resource allocation is subject to the total number of physical resource blocks (PRBs). The algorithm is derived in such a way to schedule URLLC packets first. When scheduling URLLC users, high priority is given to pending HARQ re-transmissions. Then, the buffered URLLC packets are scheduled according to their priority. Higher priority values are assigned to URLLC packets which have low latency budget. If there are PRBs left after scheduling URLLC packets, eMBB packets are scheduled.

In [8], a novel method to coexist eMBB and URLLC slices in cloud - radio access networks (C-RANs) has proposed. Two different network slices for eMBB users and URLLC users is considered. The achievable rate of eMBB user is calculated using Shannon capacity. Two different bandwidth portions for eMBB slice and URLLC slice are taken to guarantee inter-slice interference isolation. In URLLC slice, frequency division duplex (FDD) is considered. Thus, there is no interference between URLLC users. The approximation for Shannon capacity in short block length regime is used to calculate the achievable rate for URLLC users. The optimization problem is formulated to minimize the total power consumption under total bandwidth constraint, latency constraints of URLLC users and total power constraint of the base station.

Usually, in 5G networks, the time domain is divided into 1 ms time slots. The authors of [3] have mentioned that in general, resource blocks are allocated to the eMBB users at the beginning of the time slot. If URLLC user arrives in the middle of time slot which already allocated to an eMBB user, it cannot be delayed until the next time slot because of the low latency budget of URLLC users. 5G NR has a flexible frame structure which is not available in 4G, i.e. in 5G a time slot can divide into mini-slots. Usually, a time slot is consists of 14 orthogonal frequency division multiplexing (OFDM) symbols, but a mini-slot has a variable length. Mini-slot can have 2, 3, 4, ... symbols [24]. Figure 2.3 shows the difference between time slot and mini-slot.

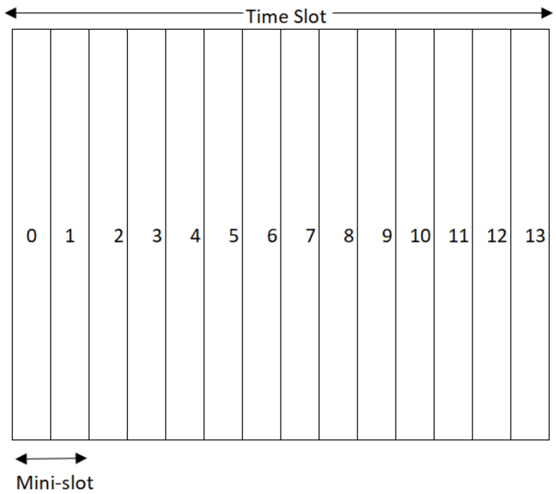

Figure 2.3. Illustration of the difference between time slot and mini-slot.

Authors in [3] have suggested to use next available mini-slot for the incoming URLLC user immediately. A system that the resource block is first allocated to eMBB users at the beginning of the time slot is proposed. When a new URLLC user arrives in a middle of a time slot while the resource block already allocates to an eMBB user, the resources are assigned to the URLLC user according to an optimization problem that maximizes the data rate of eMBB users subject to URLLC reliability constraints. 2-dimensions hopifid neural networks is used to solve the above-mentioned optimization problem.

Furthermore, the flexible frame structure of the 5G NR is used for the resource scheduling in the coexistence of eMBB and URLLC users in [25]. The concept of time slot and mini-slot is used for resource allocation. The resource scheduling strategy has two phases. In the first phase, the optimization problem is solved to maximize the throughput of eMBB users in a way such that reliability requirements of URLLC users satisfied. Resources to URLLC users are allocated by puncturing the resources of eMBB users. In the second phase, resources are allocated to eMBB users whose resources are punctured in the first phase.

2.6 Admission control

The admission control concept can interpret in two ways. Admission control in wireless networks can be interpreted as finding the maximum amount of traffic or maximum number of users that can be admitted simultaneously to the system while efficiently using the available resources and satisfying QoS requirements. Another interpretation for admission control is deciding the newly arriving traffic or newly arriving user can be admitted to the system with the available resources of the system and the QoS requirements of arriving traffic. Most of the literature on admission control are talking about finding maximum number of users that can be admitted simultaneously to the system while efficiently using the available resources and satisfying QoS requirements. The problem of joint admission control and finding optimal beamformer is an NP-hard problem [11, 12]. It is needed to search for an algorithm which can solve this problem.

When we apply admission control for the multi-user downlink network, mathematically the optimization problem can be described in two stages. In the first stage of the problem, the subset of users that can be admitted is found as follows [9]:

$$\begin{aligned} &\mathcal{U}_o = \arg\max\nolimits_{\mathcal{U}\subseteq\{1,\cdots,K\}} |\mathcal{U}| \\ \text{subject to}\quad &\frac{|\mathbf{h}_k^H \mathbf{m}_k|^2}{\sum_{i\neq k,\, i\in\mathcal{U}} |\mathbf{h}_k^H \mathbf{m}_i|^2 + N_0} \geq \gamma_k^{target}, \forall k \in \mathcal{U} \qquad (17a) \\ &\sum_{k\in\mathcal{U}} \parallel \mathbf{m}_k \parallel_2^2 \leq P_{total}, \qquad (17b) \end{aligned}$$

where $|\mathcal{U}|$ is the cardinality of $\mathcal{U}$, $\mathcal{U}$ is the set of users, $\mathbf{h}_k \in \mathbb{C}^T$ is the channel vector of kth user, $\mathbf{m}_k \in \mathbb{C}^T$ is the beamforming vector of kth user, γ_k^{target} is the minimum SINR requirement for kth user, P_{total} is the maximum available power in the base station and N_0 is the single-sided noise spectral density. From the stage one, it can be found the set $\mathcal{U}_o$, which is the set of admitted users.

In the second stage, optimal beamforming vectors are founded from the set of admitted users ($\mathcal{U}_o$), and the optimization problem is as follows [9]:

$$\begin{aligned} \text{minimize}\quad &\sum\nolimits_{k\in\mathcal{U}_o} \parallel \mathbf{m}_k \parallel_2^2 \\ \text{subject to}\quad &\frac{|\mathbf{h}_k^H \mathbf{m}_k|^2}{\sum_{i\neq k,\, i\in\mathcal{U}_o} |\mathbf{h}_k^H \mathbf{m}_i|^2 + N_0} \geq \gamma_k^{target}, \forall k \in \mathcal{U}_o \qquad (18a) \\ &\sum_{k\in\mathcal{U}_o} \parallel \mathbf{m}_k \parallel_2^2 \leq P_{total}, \qquad (18b) \end{aligned}$$

where the optimization variables are $m_k \in \mathbb{C}^T, \forall k \in \mathcal{U}_o$.

Authors in [9] have mentioned that admission control is an important area in multi-user downlink networks where there are many co-channel users and services. A single-stage reformulation to solve the admission control problem which is mentioned in optimization problems (17) and (18). Two algorithms to solve the single-stage optimization problem is proposed in [9]. In the first algorithm, semi-definite relaxation approach is used. In the second algorithm, the second-order cone programming approach is used.

In the second approach, the optimization problem is formulated for admission control using second order cone programming as follows [9]:

$$\begin{aligned} \text{minimize} \quad & \textstyle\sum_{k=1}^{K} \| \mathbf{m}_k \|_2^2 + M \sum_{k=1}^{K} s_k^2 \\ \text{subject to} \quad & \mathbf{h}_k^H \mathbf{m}_k + s_k \geq \sqrt{\gamma_k^{target} \textstyle\sum_{i=1, i\neq k}^{K} |\mathbf{h}_k^H \mathbf{m}_i|^2 + \gamma_k^{target} N_0}, \; k = 1, \ldots, K \quad (19a) \\ & \textstyle\sum_{k=1}^{K} \| \mathbf{m}_k \|_2^2 \leq P_{total} \quad (19b) \\ & \mathrm{Im}(\mathbf{h}_k^H \mathbf{m}_k) = 0, \forall k, \quad (19c) \end{aligned}$$

where M is a large positive constant, s_k is the auxiliary variable of kth user and optimization variables are $\mathbf{m}_k \in \mathbb{C}^T$ for $k = 1, \ldots, K$ and $s_k \forall k$. Authors in [9] are suggesting to solve the problem with sufficiently large M and admit the users who have a small value for the auxiliary variable s_k [9].

Authors in [10] are saying that optimal bandwidth allocation and optimal power allocation is better than equal bandwidth allocation and equal power allocation for the network where there are limited bandwidth and limited power resources. Suboptimal greedy search algorithm is suggested to solve the admission control problem to find optimal power and bandwidth allocation.

Multi-user admission control and beamformer optimization for the MISO heterogeneous networks is considered in [9]. The admission control problem for the network in which femtocells and macrocells coexist is considered. The optimization problem is formulated to find maximum number of femtocell users that can be admitted while satisfying SINR constraints of both femtocell users and macrocell users and to find optimal beamformers. This problem is formulated in two stages. In the first stage, the possible number of femtocell users is found. In the second stage, optimal beamformers for macrocell users and admitted femtocell users are found. Authors in [9] have introduced three schemes to solve the above-mentioned optimization problem.

In scheme 1, a method with full coordination between macrocell base stations and femtocell base stations is introduced. Inflation based and deflation based algorithms are adopted to get suboptimal solutions for the above-mentioned problem. Inflation based algorithm is based on adding femtocell users until the problem becomes infeasible. In deflation based algorithm all users are added first and dropped users until the problem becomes feasible. Scheme 1 is a centralized approach.

In scheme 2, a method with limited coordination between femtocell base stations and macrocell base stations is introduced. In this approach first the amount of interference from femtocell users to macrocell users which can be tolerated by macrocell users is calculated. Then the femtocell users are admitted in a way such that the interference to macrocell users does not increase the pre-calculated tolerable interference value.

In scheme 3, a method with no coordination between femtocell base stations and macrocell base stations is introduced. In this approach, zero-forcing precoding is used to femtocell users to suppress interference from femtocell users to macrocell users and perform admission control autonomously at each femtocell base station. Femtocell base stations and macrocell base stations individually do the power allocation and optimal beamformer allocation.

The problem of joint admission control and beamforming (JACoB) for a coordinated multi-cell MISO downlink network has considered in [26]. JACoB is formulated to maximize the admitted number of users whose QoS requirements have satisfied. The solutions are derived using block coordinate decent method. Therefore the algorithm is decentralized and can easily decompose for per base station.

The problem of admission control in a multi-cell downlink MISO system has been studied in [12]. The admission control problem is formulated as an ℓ_0 minimization problem. Authors in [12] have proposed two algorithms, a centralized algorithm and a distributed algorithm to solve the problem of admission control. The centralized algorithm is derived using sequential convex programming. Furthermore, the distributed algorithm is derived using the consensus-based alternating direction method of multipliers (ADMM).

Most of the literature have used deflation based algorithm to solve the problem in admission control in a wireless network. Deflation based algorithm drops users in every iteration until optimal number of possible admitted users found [11]. However, the authors in [12] have introduced a centralized algorithm which finds the possible number of admitted users at the convergence. It has been shown that the centralized algorithm in [12] has lower execution time compared with deflation based algorithms. Therefore we adopt the centralized algorithm in [12] when we derive a solution to our problem.

2.7 Sequential convex programming

Sequential convex programming is a method to find local optimal solutions for the non-convex problems [27]. In this method result is based on the initial point. Most of the time, this method is not possible to end up with an optimal solution. Sometimes it provides a suboptimal solutions. Let us consider a non-convex problem as follows:

$$
\begin{aligned}
\text{minimize} \quad & f_0(x) \\
\text{subject to} \quad & f_i(x) \leq 0, \ i = 1, \ldots, m && \text{(20a)} \\
& h_i(x) = 0, \ i = 1, \ldots, p, && \text{(20b)}
\end{aligned}
$$

where the optimization variable is $x \in \mathbb{R}^n$. To solve this problem we take x^k as the estimate for the optimal solution, τ^k is convex trust region, $\hat{f_i}$ is the convex approximation for f_i in trust region and $\hat{h_i}$ is the convex approximation for h_i in trust region. We can obtain the optimal solution x^{k+1} by solving the following problem [27]:

$$
\begin{aligned}
\text{minimize} \quad & \hat{f_0(x)} \\
\text{subject to} \quad & \hat{f_i(x)} \leq 0, \ i = 1, \ldots, m && \text{(21a)} \\
& \hat{h_i(x)} = 0, \ i = 1, \ldots, p && \text{(21b)} \\
& x \in \tau^k. && \text{(21c)}
\end{aligned}
$$

Definition for trust region is as follows [27]:

$$
\tau^k = \{x | |x_i - x_i^k| \leq \rho_i, \ i = 1, \ldots, n\}, \tag{22}
$$

where ρ_i is predefined threshold.

2.7.1 Difference of convex programming

Let us consider an optimization problem as follows:

$$\begin{aligned} \text{minimize} \quad & f_0(x) - g_0(x) \\ \text{subject to} \quad & f_i(x) - g_i(x) \le 0, \ i = 1, \ldots, m, \end{aligned} \tag{23a}$$

where the optimization variable is $x \in \mathbb{R}^n$. This problem is known as difference of convex programming. $f_i - g_i$ is known as difference of convex function [27].

2.7.2 Convex-Concave procedure

Let us consider, f_i and g_i in problem (23) are differentiable functions. To solve this kind of optimization problem we can use the first order approximation for $f(x) - g(x)$ as follows [27]:

$$f(\hat{x}) = f(x) - (g(x^k) + \nabla g(x^k)^T (x - x^k)). \tag{24}$$

The procedure of solving difference convex problems using sequential convex programming is known as the convex-concave procedure. First, we need to decide on an initial feasible point. Then we need to solve the problem,

$$\begin{aligned} \text{minimize} \quad & f_0(x) - (g_0(x^k) + \nabla g_i(x^k)^T (x - x^k)) \\ \text{subject to} \quad & f_i(x) - (g_i(x^k) + \nabla g_i(x^k)^T (x - x^k)) \le 0, \ i = 1, \ldots, m, \end{aligned} \tag{25a}$$

by iterating until the difference between the objective value of two consecutive interactions is less than some predefined threshold [28].

3 SYSTEM MODEL AND PROBLEM FORMULATION

In this chapter, we formulate the problem of admission control for eMBB users in the coexistence of eMBB and URLLC in a 5G network. Also in this chapter we present the solution approach and an algorithm to find a suboptimal solution.

As we mentioned in Chapter 2, existing literature shows that there is a need to coexist eMBB and URLLC users and there are methods and algorithms to jointly schedule eMBB and URLLC users in one resource [3, 6–8]. Also there are existing literature which intend to find a solution for admission control problem in wireless network [9–12]. It is useful to find a way to do admission control in networks in which eMBB and URLLC users coexist which is not addressed in the existing literature.

We suggest orthogonal spectrum sharing between eMBB and URLLC users to coexist them. We propose an algorithm to find a maximum number of eMBB users who have sufficient data rates that can be admitted to the system while allocating power, bandwidth and beamforming directions to all URLLC users whose latency and reliability requirements are always guaranteed. The objective of the problem is to maximize the number of admitted eMBB users subject to following constraints.

- Signal-to-interference-plus-noise ratio constraint for eMBB users
- Signal-to-noise ratio constraint for URLLC users in order to satisfy high reliability and low latency requirements of URLLC users
- Transmit power constraint of the base station
- Total bandwidth constraint of the system

We have formulated this problem as an ℓ_0 minimization problem. We have used sequential convex programming method to solve the problem, since it is an NP-hard problem [27, 28]. SINR constraint for eMBB user is formulated using the achievable rate of eMBB users which is derived through Shannon's capacity. Since URLLC has short packet length we have used the approximation for Shannon's capacity in finite blocklength regime to derive the maximum achievable rate for URLLC users [29]. SNR constraint for URLLC users has formulated using the concept of effective bandwidth in a way such that guaranteeing the latency and reliability requirements of URLLC users [5].

3.1 System model

In this section we present the system model. We consider the downlink of a single-cell MISO system. We assume that the base station have T transmit antennas. The set of users are denoted by $\mathcal{U}$. The set of all eMBB users denoted by $\mathcal{U}_e \subset \mathcal{U}$ and they are labeled with the integer values $k = 1, \ldots, K$. We use the notation $\mathcal{U}_u \subset \mathcal{U}$ to denote the set of all URLLC users and they are labeled with the integer values $j = 1, \ldots, J$. We assume that all users have only one receive antenna. Figure 3.1 shows an illustration of the system model.

We propose orthogonal spectrum sharing between eMBB and URLLC users to coexist them. Let the total bandwidth of the system is B_{total}, total bandwidth for eMBB users

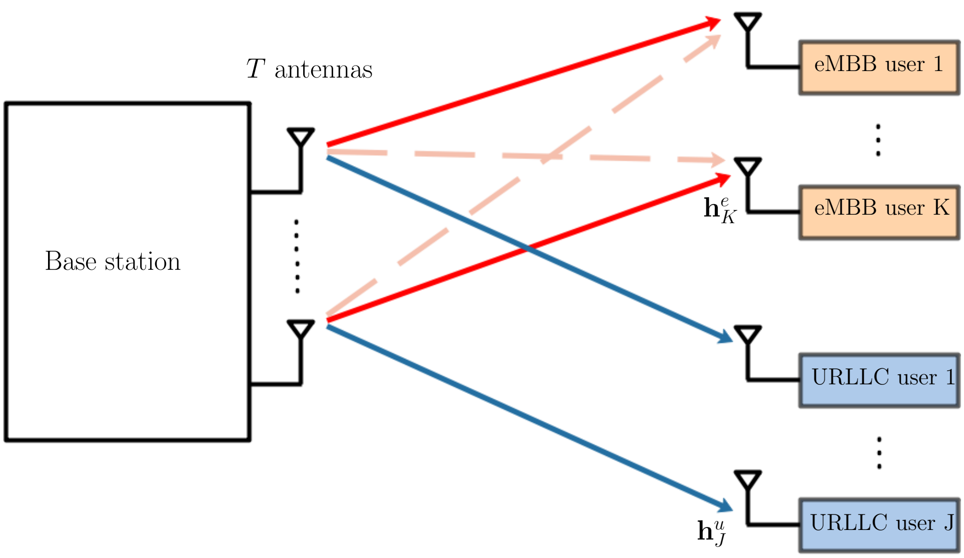

Figure 3.1. Illustration of the system model.

is b^e and total bandwidth for URLLC users is b^u. Thus, the expression for bandwidth sharing can be written as

$$B_{total} = b^e + b^u. \tag{26}$$

There is no interference between eMBB users and URLLC users since they are getting a separate portions of bandwidth. Furthermore, we suggest to give a separate portions of bandwidth for each URLLC user. Moreover, we consider orthogonal frequency division multiple access (OFDMA) for URLLC users. This implies URLLC users are allocated orthogonal resources, and hence there is no interference between URLLC users.

3.2 Problem formulation

In this section we present the problem of admission control. The signal vector transmitted by the base station is given by

$$\mathbf{X} = \sum_{k \in \mathcal{U}} \mathbf{m}_k d_k, \tag{27}$$

where d_k is the normalized data symbol of the kth user, and we assume that the data streams are independent. The beamforming vector $\mathbf{m}_k \in \mathbb{C}^T$ can be written as

$$\mathbf{m}_k = \sqrt{p_k}\mathbf{u}_k, \tag{28}$$

where $\mathbf{u}_k \in \mathbb{C}^T$ is the normalized beamformer, $\mathbf{m}_k \in \mathbb{C}^T$ is the beamforming vector of the kth user and p_k is the power of kth user.

Received signal vector of the kth eMBB user is given by

$$y_k^e = (\mathbf{h}_k^e)^H \mathbf{m}_k^e d_k^e + \sum\nolimits_{i=1, i \neq k}^{K} (\mathbf{h}_k^e)^H \mathbf{m}_i^e d_i^e + w_k^e, \tag{29}$$

where $\mathbf{h}_k^e \in \mathbb{C}^T$ is the channel vector from base station to eMBB user k, $\mathbf{m}_k^e \in \mathbb{C}^T$ is the beamforming vector of the kth eMBB user and $w_k^e \sim CN(0, \sigma_e^2)$ is the additive white Gaussian noise (AWGN) at eMBB user k. We take the noise variance as $\sigma_e^2 = N_0 b^e$, where N_0 is single-sided noise spectral density. Since the eMBB users and URLLC users are in two portions of bandwidth there is no interference between URLLC users and eMBB users.

In general, URLLC has an end to end delay less than 1 ms. Therefore, the channel coherence time is greater than the end to end delay. This means URLLC users have a quasi-static channel and the rate of URLLC users can take as a constant for a given resource allocation policy [5, 19–21]. Received signal of the jth URLLC user can be written as

$$y_j^u = (\mathbf{h}_j^u)^H \mathbf{m}_j^u d_j^u + w_j^u, \tag{30}$$

where $\mathbf{h}_j^u \in \mathbb{C}^T$ is the channel vector from base station to URLLC user j, $\mathbf{m}_j^u \in \mathbb{C}^T$ is the beamforming vector of the jth URLLC user and $w_j^u \sim CN(0, \sigma_{j,u}^2)$ is the additive white Gaussian noise (AWGN) at URLLC user j. We take the noise variance as $\sigma_{j,u}^2 = N_0 b_j^u$, where b_j^u is the bandwidth allocated to the jth URLLC user. Since the URLLC users are allocated orthogonal resources, there is no interference between URLLC users.

The received SINR of kth eMBB user can be expressed as

$$\gamma_k^e = \frac{|(\mathbf{h}_k^e)^H \mathbf{m}_k^e|^2}{\sum_{i=1, i\neq k}^{K} |(\mathbf{h}_k^e)^H \mathbf{m}_i^e|^2 + N_0 b^e}. \tag{31}$$

The received SINR of jth URLLC user can be expressed as

$$\gamma_j^u = \frac{|(\mathbf{h}_j^u)^H \mathbf{m}_j^u|^2}{N_0 b_j^u}. \tag{32}$$

The maximum achievable rate for kth eMBB user can be written as

$$R_k^e = b^e \log_2(1 + \gamma_k^e). \tag{33}$$

We assume that the target rate for an eMBB user is R_{target}. Thus, the target SINR for the kth eMBB user can be expressed as

$$\gamma_k^{e,th} = 2^{\frac{R_{target}}{b_e}} - 1. \tag{34}$$

The target rate for eMBB users can be achieved if its SINR is greater than the SINR threshold, $\gamma_k^{e,th}$, i.e.,

$$\gamma_k^e \geq \gamma_k^{e,th}. \tag{35}$$

We consider that the maximum packet delay threshold, D_{max} is 1 ms and overall reliability requirement, ϵ is 1×10^{-5}. The overall reliability is the overall packet loss probability of a single user which is the combination of transmission error probability and queuing-delay violation probability. The overall reliability, ϵ can be expressed as

$$\epsilon = \epsilon_c + \epsilon_q, \tag{36}$$

where ϵ_c is the transmission-error probability and ϵ_q is the queuing-delay violation probability.

Furthermore, we assume that downlink transmissions only requires one frame and duration of one frame is T_f. Furthermore, the latency of the backhaul is T_f. Thus, we can obtain end to end queuing delay as follows [19]:

$$D_q = D_{max} - 2T_f. \tag{37}$$

If channel state information (CSI) is known at the transmitter and receiver, in quasi-static, interference-free, flat fading channel, the maximum achievable rate of the jth user can be approximated as [19]

$$R_j^u = \frac{\tau b_j^u}{\ln 2}[\log_2(1+\gamma_j^u) - \sqrt{\frac{V_j^u}{\tau b_j^u}} Q^{-1}(\epsilon_c)] \ bits/frame, \tag{38}$$

where τ is duration for data transmission in one frame, Q^{-1} is the inverse Q function and V_j^u is channel dispersion of URLLC user j, which is given by [19]

$$V_j^u = 1 - \frac{1}{(1+\gamma_j^u)^2}. \tag{39}$$

Therefore, the queuing delay requirements (D_q and ϵ_q) can be satisfied when the achievable rate is greater than or equal to the effective bandwidth [5,19–21]. The effective bandwidth for a Poisson process with arrival packet rate λ, can be expressed as [5]

$$E^B = \frac{\mu T_f \ln \frac{1}{\epsilon_q}}{D_q \ln(\frac{T_f \ln \frac{1}{\epsilon_q}}{\lambda D_q} + 1)} \ bits/frame, \tag{40}$$

where μ is the number of bits contained in each packet. We can obtain the SNR required to satisfy queuing delay requirements by taking $R_j^u = E^B$ and by substituting $V_j^u \approx 1$ to achieve the lower bound. Thus, the threshold for SNR of URLLC user j is given by

$$\gamma_j^{u,th} = \exp[\frac{E^B \ln 2}{\tau b_j^u} + \sqrt{\frac{1}{\tau b_j^u}} Q^{-1}(\epsilon_c)] - 1. \tag{41}$$

Latency and reliability requirements of jth URLLC user is satisfied if SNR of the jth URLLC user is greater than the SINR threshold $\gamma_j^{u,th}$, i.e.,

$$\gamma_j^u \geq \gamma_j^{u,th}. \tag{42}$$

We assume that the power allocation for both eMBB and URLLC users is less than or equal to maximum transmit power at the base station P_{total}, i.e.,

$$\sum_{k=1}^{K} \| \mathbf{m}_k^e \|_2^2 + \sum_{j=1}^{J} \| \mathbf{m}_j^u \|_2^2 \leq P_{total}. \tag{43}$$

Furthermore, we assume that bandwidth allocation for all eMBB users and for each URLLC user is less than or equal to the total bandwidth of the system B_{total}, i.e.,

$$b^e + \sum_{j=1}^{J} b_j^u \leq B_{total}. \tag{44}$$

We consider the admission of eMBB users who have satisfied the target rate while allocating power, bandwidth and beamforming directions to all URLLC users who have satisfied the latency and reliability requirements under the power and bandwidth constraints.

Our goal is to maximize the number of admitted eMBB users such that all the constraints related the eMBB and URLLC users are satisfied. Thus, we need to maximize the sum of the cardinalities of $\mathcal{U}_e$. To formulate this problem as a mathematical optimization problem we define the non negative auxiliary variable s_k and relax the SINR constraint for the kth eMBB user as follows:

$$\gamma_k^e \geq \gamma_k^{e,th} - s_k. \tag{45}$$

In the Equation (45), we can obtain the Equation (35) when $s_k = 0$. That means when $s_k = 0$ the SINR constraint of kth eMBB user is satisfied. Therefore, in order to maximize the number of admitted eMBB users who achieve the target rate, we have to minimize the number of users that require a strictly positive value of auxiliary variable s_k [12]. In other words we have to increase number of times when $s_k = 0$. It can be achieved by minimizing ℓ_0 norm of the vector consists of auxiliary variables. Hence the optimization problem of admission control for eMBB in the coexistence of URLLC and eMBB users can be expressed as follows:

$$\begin{aligned}
\text{minimize} \quad & \| \mathbf{s} \|_0 \\
\text{subject to} \quad & \gamma_k^e \geq \gamma_k^{e,th} - s_k, \; k = 1, \ldots, K && \text{(46a)} \\
& \gamma_j^u \geq \gamma_j^{u,th}, \; j = 1, \ldots, J && \text{(46b)} \\
& b^e + \sum\nolimits_{j=1}^{J} b_j^u \leq B_{total} && \text{(46c)} \\
& b^e \geq 0 && \text{(46d)} \\
& b_j^u \geq 0, \; j = 1, \ldots, J && \text{(46e)} \\
& \sum\nolimits_{k=1}^{K} \| \mathbf{m}_k^e \|_2^2 + \sum\nolimits_{j=1}^{J} \| \mathbf{m}_j^u \|_2^2 \leq P_{total} && \text{(46f)} \\
& s_k \geq 0, \; k = 1, \ldots, K, && \text{(46g)}
\end{aligned}$$

where $\mathbf{s} = [s_1, ..., s_k]^T$ and optimization variables are $\{s_k, \mathbf{m}_k^e\}$ for $k = 1, \ldots, K$, b^e and $\{\mathbf{m}_j^u, b_j^u\}$ for $j = 1, \ldots, J$.

3.3 Algorithm derivation

In this section we present the way we derived the algorithm to solve the optimization problem (46). Problem (46) has an ℓ_0 objective function and it is known as an NP-hard problem. Thus, it is exponentially complex to find an optimal solution to this problem. Therefore, we provide a suboptimal algorithm that can find a suboptimal solution to the problem [12]. Thus, we derive an algorithm to solve the problem using ℓ_0 approximation method and sequential convex programming [12]. We approximate the objective function with a concave function $\sum_{k=1}^{K} log(s_k + \delta)$ where δ is small positive constant and $s_k \geq 0$, $k = 1, \ldots, K$ [12]. We denote the interference plus noise experienced by the kth eMBB user, by the variable β_k as follows:

$$\beta_k = \sum\nolimits_{i=1, i \neq k}^{K} |(\mathbf{h}_k^e)^H \mathbf{m}_i^e|^2 + N_0 b^e. \tag{47}$$

Approximation for the optimization problem (46) is as follows:

$$\begin{aligned}
\text{minimize} \quad & \textstyle\sum_{k=1}^{K} log(s_k+\delta) \\
\text{subject to} \quad & 2^{\frac{R_{target}}{b^e}} - 1 - s_k - \frac{|(\mathbf{h}_k^e)^H \mathbf{m}_k^e|^2}{\beta_k} \leq 0, \;\; k=1,\ldots,K && \text{(48a)} \\
& \textstyle\sum_{i=1, i\neq k}^{K} |(\mathbf{h}_k^e)^H \mathbf{m}_i^e|^2 + N_0 b^e \leq \beta_k, \;\; k=1,\ldots,K && \text{(48b)} \\
& \exp\Big[\frac{E^B \ln 2}{\tau b_j^u} + \sqrt{\frac{1}{\tau b_j^u}} Q^{-1}(\epsilon_c)\Big] - 1 - \frac{|(\mathbf{h}_j^u)^H \mathbf{m}_j^u|^2}{N_0 b_j^u} \leq 0, \;\; j=1,\ldots,J && \text{(48c)} \\
& \text{constraints (46c), (46d), (46e), (46f), (46g),} && \text{(48d)}
\end{aligned}$$

where optimization variables are $\{s_k,\ \mathbf{m}_k^e\}$ for $k=1,\ldots,K$, b^e and $\{\mathbf{m}_j^u,\ b_j^u\}$ for $j=1,\ldots,J$.

Problem (48) is still non convex, because still it has a concave objective function and non convex constraint functions, i.e., the constraints (48a) and (48c) are non convex. Therefore, to solve the the problem (48), we apply sequential convex programming [12].

We denote the objective function of the problem (48) by $f(s)=\sum_{k=1}^{K} log(s_k+\delta)$. Since $f(s)$ is a concave function, we take its first order approximation, and approximation for the objective function can be denote by [12]

$$\hat{f}(\mathbf{s}) = f(\hat{\mathbf{s}}) + \sum_{k=1}^{K} \frac{(s_k - \hat{s}_k)}{(\hat{s}_k + \delta)}, \tag{49}$$

and it is evaluated at the point $\hat{\mathbf{s}}=[\hat{s}_1, ..., \hat{s}_K]$.

The constraint (48a) is in the form of difference of convex function. We apply convex - concave procedure to make the constraint (48a) convex. We define $g_k(\mathbf{m}_k^e, \beta_k)$ as $g_k(\mathbf{m}_k^e, \beta_k) = |(\mathbf{h}_k^e)^H \mathbf{m}_k^e|^2/\beta_k$. First order approximation of $g_k(\mathbf{m}_k^e, \beta_k)$ is as follows:

$$\hat{g}_k(\mathbf{m}_k^e, \beta_k) = g_k(\hat{\mathbf{m}}_k^e, \hat{\beta}_k) + \nabla g_k(\hat{\mathbf{m}}_k^e, \hat{\beta}_k)^T ((\mathbf{m}_k^e, \beta_k) - (\hat{\mathbf{m}}_k^e, \hat{\beta}_k)), \tag{50}$$

where $\nabla g_k(\hat{\mathbf{m}}_k^e, \hat{\beta}_k)$ is the gradient of $g_k(\mathbf{m}_k^e, \beta_k)$ which is evaluated at the point $(\hat{\mathbf{m}}_k^e, \hat{\beta}_k)$. $\nabla g_k(\hat{\mathbf{m}}_k^e, \hat{\beta}_k)$ is given by

$$\nabla g_k(\hat{\mathbf{m}}_k^e, \hat{\beta}_k) = \Big(\frac{2\mathbf{h}_k^e (\mathbf{h}_k^e)^H \hat{\mathbf{m}}_k^e}{\hat{\beta}_k}, \frac{-(\hat{\mathbf{m}}_k^e)^H \mathbf{h}_k^e (\mathbf{h}_k^e)^H \hat{\mathbf{m}}_k^e}{\hat{\beta}_k^2}\Big). \tag{51}$$

The constraint (48c) is also in the form of difference of convex function. We apply convex - concave procedure to make the constraint (48c) convex. We define $z_j(\mathbf{m}_j^u, b_j^u)$ as, $z_j(\mathbf{m}_j^u, b_j^u) = \frac{|(\mathbf{h}_j^u)^H \mathbf{m}_j^u|^2}{N_0 b_j^u}$. The first order approximation of $z_j(\mathbf{m}_j^u, b_j^u)$ is as follows:

$$\hat{z}_j(\mathbf{m}_j^u, b_j^u) = z_j(\hat{\mathbf{m}}_j^u, b_j^u) + \nabla z_j(\hat{\mathbf{m}}_j^u, \hat{b}_j^u)^T ((\mathbf{m}_j^u, b_j^u) - (\hat{\mathbf{m}}_j^u, \hat{b}_j^u)), \tag{52}$$

where $\nabla z_j(\hat{\mathbf{m}}_j^u, \hat{b}_j^u)$ is the gradient of $z_j(\mathbf{m}_j^u, b_j^u)$ which is evaluated at the point $(\hat{\mathbf{m}}_j^u, \hat{b}_j^u)$. $\nabla z_j(\hat{\mathbf{m}}_j^u, \hat{b}_j^u)$ is given by

$$\nabla z_j(\hat{\mathbf{m}}_j^u, \hat{b}_j^u) = \Big(\frac{2\mathbf{h}_j^u (\mathbf{h}_j^u)^H \hat{\mathbf{m}}_j^u}{N_0 \hat{b}_j^u}, \frac{-(\hat{\mathbf{m}}_j^u)^H \mathbf{h}_j^u (\mathbf{h}_j^u)^H \hat{\mathbf{m}}_j^u}{N_0 (\hat{b}_j^u)^2}\Big). \tag{53}$$

Now we approximate the optimization problem (48) using the Equations (49), (50) and (52). Approximated optimization problem is as follows:

$$\begin{aligned}
&\text{minimize} \quad \textstyle\sum_{k=1}^{K} \frac{s_k}{(\hat{s}_k+\delta)} \\
&\text{subject to} \quad 2^{\frac{R_{target}}{b^e}} - 1 - s_k - \hat{g}_k(\mathbf{m}_k^e, \beta_k) \leq 0,\ k = 1, \ldots, K && \text{(54a)} \\
&\textstyle\sum_{i=1, i\neq k}^{K} |(\mathbf{h}_k^e)^H \mathbf{m}_i^e|^2 + N_0 b^e \leq \beta_k,\ k = 1, \ldots, K && \text{(54b)} \\
&\exp\left[\frac{E^B \ln 2}{\tau b_j^u} + \sqrt{\frac{1}{\tau b_j^u}} Q^{-1}(\epsilon_c)\right] - 1 - \hat{z}_j(\mathbf{m}_j^u, b_j^u) \leq 0,\ j = 1, \ldots, J && \text{(54c)} \\
&\text{constraints (46c), (46d), (46e), (46f), (46g),} && \text{(54d)}
\end{aligned}$$

where the optimization variables are $\{s_k,\ \mathbf{m}_k^e\}$ for $k = 1, \ldots, K$, b^e and $\{\mathbf{m}_j^u,\ b_j^u\}$ for $j = 1, \ldots, J$. We have dropped the constant terms $f(\hat{\mathbf{s}})$ and $\frac{\hat{s}_k}{\hat{s}_k+\delta}$ from the objective function of problem (54), since they are not affect the solution.

3.4 Algorithm

In this section we present the proposed algorithm for the problem (54), which is summarized in *Algorithm 1.*

Algorithm 1 Algorithm for solving problem (54)

1: **initialization**: $\{s_k^0,\ (\mathbf{m}_k^e)^0,\ \beta_k^0\}$ for $k = 1, \ldots, K$, b^e and $\{(\mathbf{m}_j^u)^0,\ (b_j^u)^0\}$ for $j = 1, \ldots, J$, iteration index $p = 0$.
repeat
2: Set $\hat{\mathbf{m}}_k^e = (\mathbf{m}_k^e)^p$, $\hat{\beta}_k = \beta_k^p$ for $k = 1, \ldots, K$ and $\hat{\mathbf{m}}_j^u = (\mathbf{m}_j^u)^p$, $\hat{b}_j^u = (b_j^u)^p$ for $j = 1, \ldots, J$. Form $\hat{g}_k(\mathbf{m}_k^e, \beta_k)\ \forall k$ using the Equation (50) and $\hat{z}_j(\mathbf{m}_j^u, b_j^u)\ \forall j$ using the Equation (52).
3: Solve problem (54). Denote the solution $\{s_k^\star,\ (\mathbf{m}_k^e)^\star,\ \beta_k^\star\}$ for $k = 1, \ldots, K$ and $\{(\mathbf{m}_j^u)^\star,\ (b_j^u)^\star\}$ for $j = 1, \ldots, J$. Set $p = p + 1$.
4: Update $\{s_k^{p+1} = s_k^\star,\ (\mathbf{m}_k^e)^{p+1} = (\mathbf{m}_k^e)^\star,\ \beta_k^{p+1} = \beta_k^\star\}$ for $k = 1, \ldots, K$ and $\{(\mathbf{m}_j^u)^{p+1} = (\mathbf{m}_j^u)^\star,\ (b_j^u)^{p+1} = (b_j^u)^\star\}$ for $j = 1, \ldots, J$.
until stopping criterion is satisfied

The algorithm is iterated until the difference between the objective values of problem (54) in consecutive iterations is less than a predefined threshold.

3.5 Challenges faced during algorithm derivation

We have faced following challenges during the derivation of the algorithm. In the constraint (54a), let us define $v_1 = 2^{\frac{R_{target}}{b^e}}$ and $v_2^k = \hat{g}_k(\mathbf{m}_k^e, \beta_k)$. Then the constraint (54a) becomes

$$v_1 - 1 - s_k - v_2^k \leq 0,\ k = 1, \ldots, K \tag{55}$$

If v_1 and v_2^k both are variables solver is not able to successfully do the admission control by assigning values to the slack variable s_k. Therefore, we have to further relax the problem by making the bandwidth portion for eMBB users (b^e), constant in order to make v_1 a constant. We have achieved that by giving the ratio of eMBB and URLLC bandwidth allocation from the total bandwidth of the system as an input at the beginning of the algorithm.

4 SIMULATION AND NUMERICAL RESULTS

We simulate the proposed algorithm in order to prove the correctness and effectiveness of our algorithm. This chapter provides the simulation setup and numerical results.

4.1 Simulation setup

In our simulations, the downlink of a single-cell MISO system is considered. We assume that the base station is equipped with four transmit antennas. There are eight eMBB users and eight URLLC users in the system. To model the channel gains, we have used the exponential path loss model which is given by

$$\mathbf{h}_k = (\frac{r_k}{r_0})^{-\alpha} \mathbf{c}_k, \tag{56}$$

where $\mathbf{h}_k \in \mathbb{C}^T$ is the channel vector from base station to kth user, r_k is the distance from base station to kth user, r_0 is the far-field reference distance, α is the path loss exponent and $\mathbf{c}_k$ is small scale fading which is arbitrary chosen from circularly symmetric complex Gaussian vector distribution with mean zero and identity covariance matrix. We assume that both eMBB and URLLC users are distributed uniformly around the base station within the distance 10 m and 100 m. We take the bandwidth allocation between eMB users and URLLC users as total bandwidth for eMBB users $= B_{total} \times \frac{1}{2}$ and the total bandwidth of URLLC users $= B_{total} \times \frac{1}{2}$.

We have done the simulations using Matalb. We have solved the admission control problem using CVX with MOSEK solver. Furthermore, the simulation parameters mentioned in Table 4.1 are assumed.

Table 4.1. Simulation parameters

Far field distance r_0	1 m
Path loss exponent α	2
Overall reliability requirement ϵ	1×10^{-5}
Transmission error probability $\epsilon_c = \epsilon/2$	5×10^{-6}
Queueing-delay violation probability $\epsilon_q = \epsilon/2$	5×10^{-6}
E2E delay requirement D_{max}	1 ms
Maximum queueing delay D_q	0.8 ms
Duration of each frame T_f	0.1 ms
Duration of data transmission in one fame τ	0.05 ms
Packet size μ	20 bytes
Maximum transmit power P_{total}	33 dBm
Arrival packet rate λ	0.2 packets/frame
Single-sided noise spectral density N_0	-83.98 dBm/Hz [8]
Total bandwidth of the system B_{total}	200 MHz
Target rate for an eMBB user R_{target}	200 Mbps

The algorithm is iterated until the difference between the objective values of problem (54) in consecutive iterations is less than 0.001.

4.2 Numerical results

4.2.1 Scenario 1

We simulate an arbitrarily chosen single channel and topology realization with the simulation parameters mentioned in the Section 4.1. The objective value $f(s) = \sum_{k=1}^{K} log(s_k + \delta)$ is calculated for every iteration until convergence. s_k is the auxiliary variable for the kth eMBB user. Furthermore, we count the admitted number of eMBB users at each iteration. Then we draw the objective value versus iteration and number of admitted users versus iteration in the same graph in order to check the convergence of the algorithm.

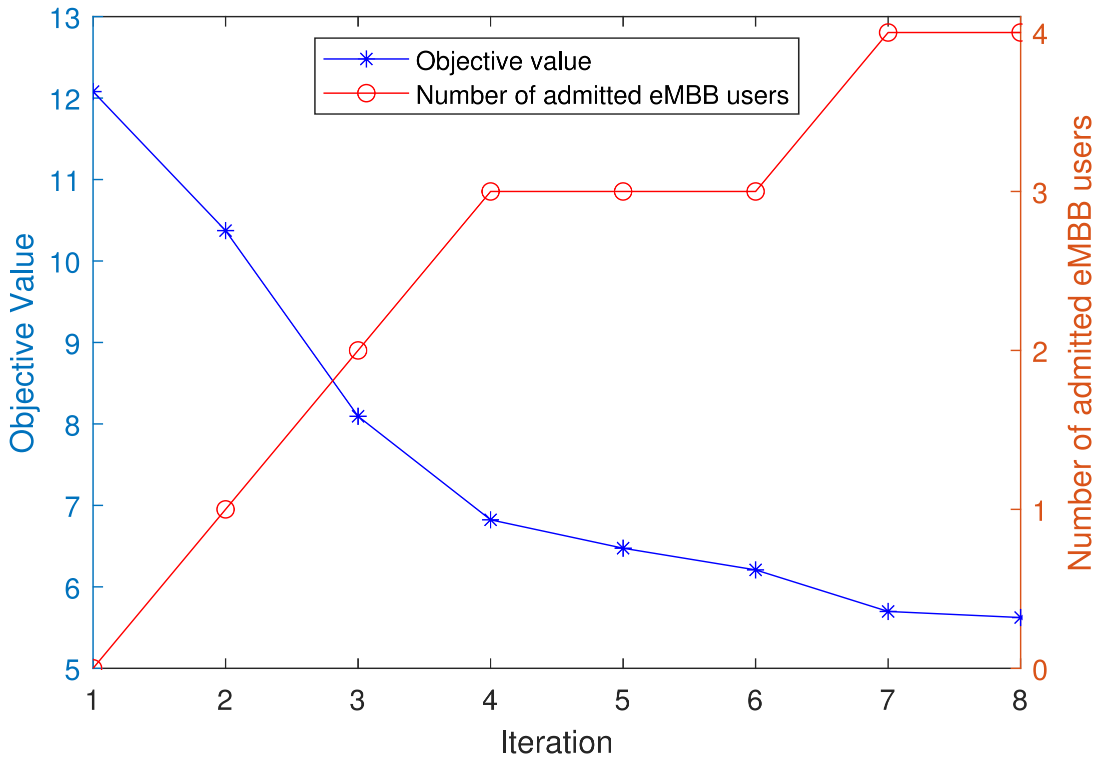

Figure 4.1. Objective value versus iterations and number of admitted users versus iterations.

Figure 4.1 shows the behaviour of the convergence of the *Algorithm 1*. According to the Figure 4.1, the objective value is minimized and converge after eight iterations. The admitted number of users is increase with iterations. At the convergence we are able to get the optimal solution of the algorithm. According to the Figure 4.1, four eMBB users can be admitted to the system with the default simulation parameters mentioned in the Section 4.1. Therefore, we can observe that, our algorithm is possible to obtain a optimal solution in few iterations.

4.2.2 Scenario 2

Then we evaluate how the admitted number of eMBB users behave with the target rate for eMBB user and the total bandwidth of the system. The algorithm has been run over 250 channel and topology realizations with the simulation parameters mentioned in Section 4.1. We simulate it for different ratios of eMBB and URLLC bandwidth allocation from the total bandwidth of the system. Table 4.2 shows the three different cases that we have simulated.

Table 4.2. eMBB and URLLC bandwidth allocation

Case	**Bandwidth portion for eMBB (b^e)**	**Bandwidth portion for URLLC (b^u)**
1	$B_{total} \times \frac{3}{4}$	$B_{total} \times \frac{1}{4}$
2	$B_{total} \times \frac{1}{2}$	$B_{total} \times \frac{1}{2}$
3	$B_{total} \times \frac{1}{4}$	$B_{total} \times \frac{3}{4}$

The variation of the admitted number of users with the target rate of eMBB users for different values of total bandwidth for case 1, case 2 and case 3 shows in Figures 4.2, 4.3 and 4.4 respectively.

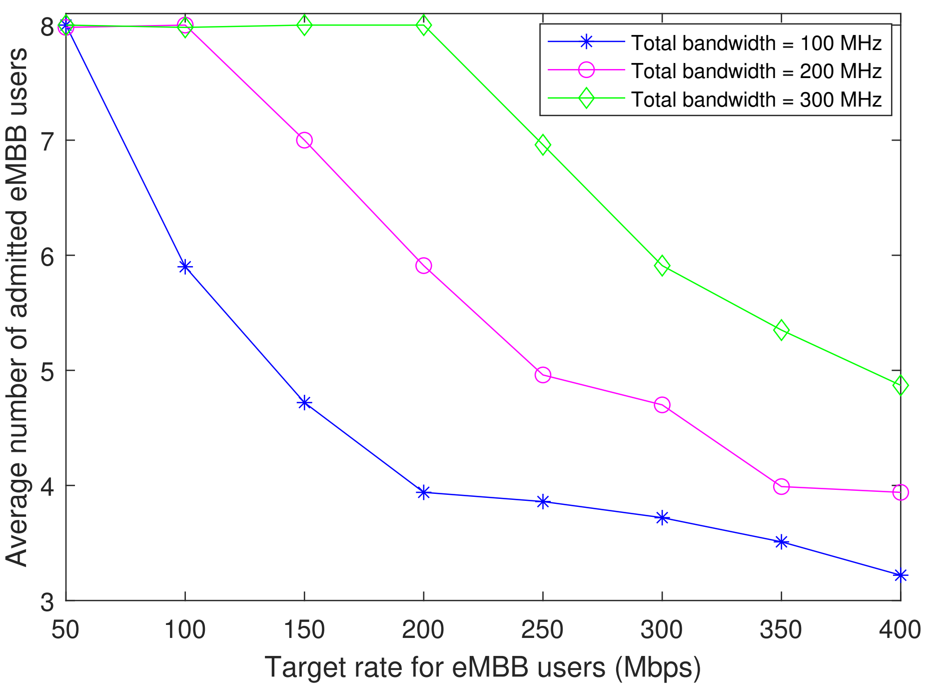

Figure 4.2. Admitted eMBB users versus target rate for eMBB users when $b^e = B_{total} \times \frac{3}{4}$, $b^u = B_{total} \times \frac{1}{4}$, and each URLLC user has variable bandwidth which is the optimal bandwidth allocation.

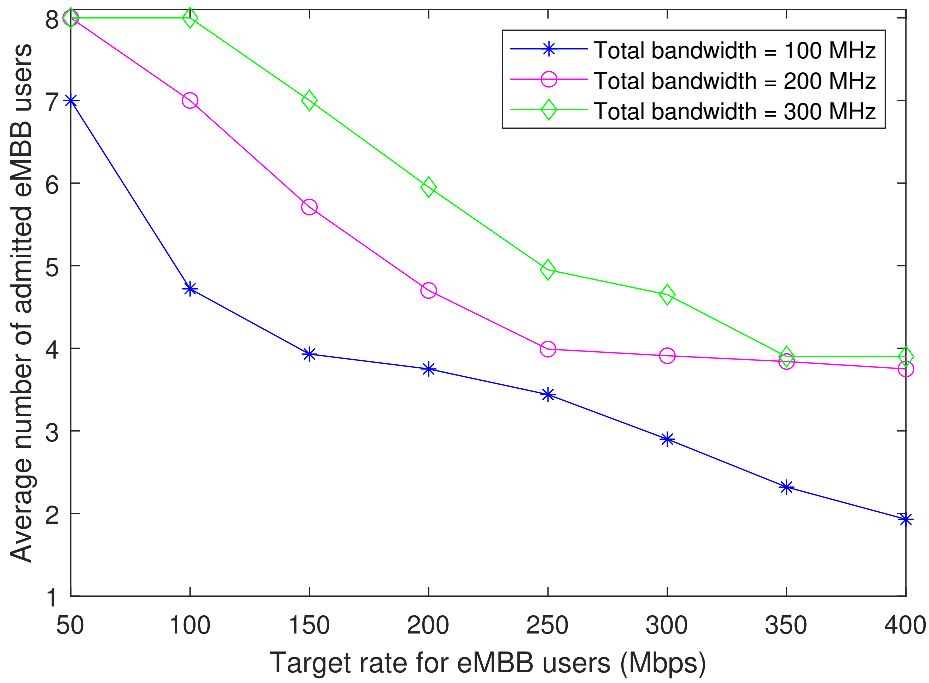

Figure 4.3. Admitted eMBB users versus target rate for eMBB users when $b^e = B_{total} \times \frac{1}{2}$, $b^u = B_{total} \times \frac{1}{2}$, and each URLLC user has variable bandwidth which is the optimal bandwidth allocation.

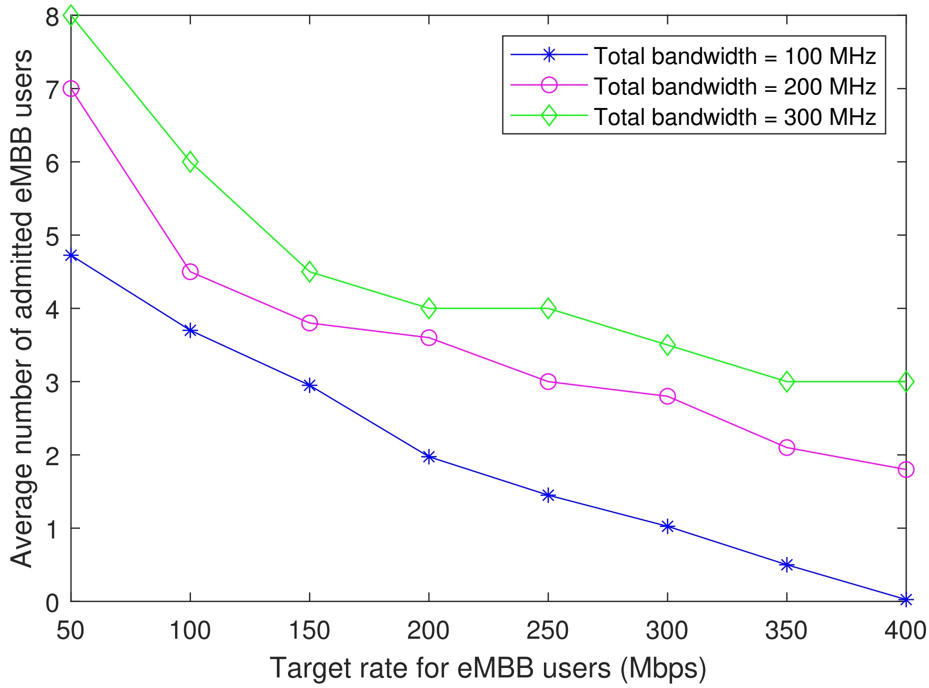

Figure 4.4. Admitted eMBB users versus target rate for eMBB users when $b^e = B_{total} \times \frac{1}{4}$, $b^u = B_{total} \times \frac{3}{4}$, and each URLLC user has variable bandwidth which is the optimal bandwidth allocation.

From the Figures 4.2, 4.3 and 4.4, we can observe that the admitted number of eMBB users is increasing with the increase of the total bandwidth of the system. Another observation is the increase of admitted number of eMBB users with the increase of bandwidth portion allocation for eMBB users. Furthermore, we can observe that the admitted number of eMBB users is decreasing with the increase in the target rate for eMBB users.

4.2.3 Scenario 3

Then we evaluate how the admitted number of eMBB users behaves with the maximum power of the base station. The algorithm has been run over 250 channel and topology realizations with the simulation parameters mentioned in Table 4.1. We take the bandwidth sharing as total bandwidth for eMBB users $= B_{total} \times \frac{1}{2}$ and the total bandwidth of URLLC users $= B_{total} \times \frac{1}{2}$. The variation of the admitted number of users with the maximum power of the base station shows in Figure 4.5 below.

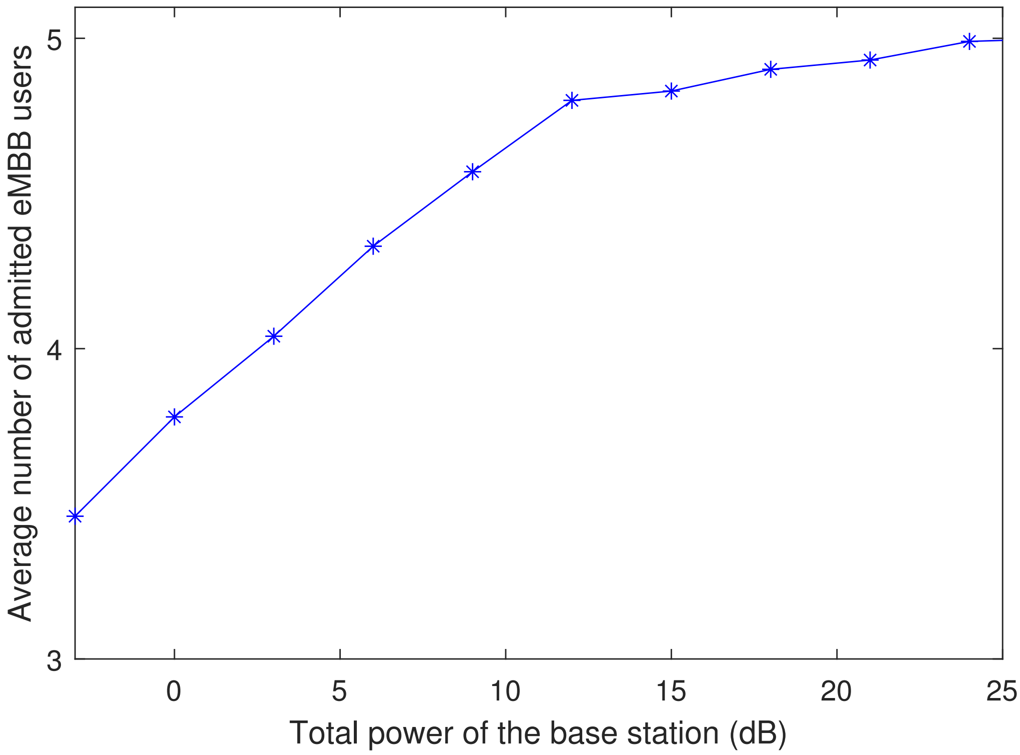

Figure 4.5. Admitted eMBB users versus maximum power of the base station.

We can observe that the admitted number of eMBB users is increasing with the increase of the maximum available power for transmission at the base station.

4.2.4 Scenario 4

In this scenario, we evaluate the impact of URLLC users on the number of admitted eMBB users in the system. The algorithm has been over 250 channel and topology realizations with the simulation parameters mentioned in Table 4.1. We evaluate how the admitted number of eMBB users change with the increase of the URLLC users in the system for two cases mentioned in Table 4.3. The variation of the admitted number of users with the number of URLLC users in the system for case 1 and case 2 illustrates in the Figure 4.6.

Table 4.3. eMBB and URLLC bandwidth allocation

Case	**Bandwidth portion for eMBB (b^e)**	**Bandwidth portion for URLLC (b^u)**
1	$B_{total} \times \frac{1}{5}$	$B_{total} \times \frac{4}{5}$
2	$B_{total} \times \frac{1}{2}$	$B_{total} \times \frac{1}{2}$

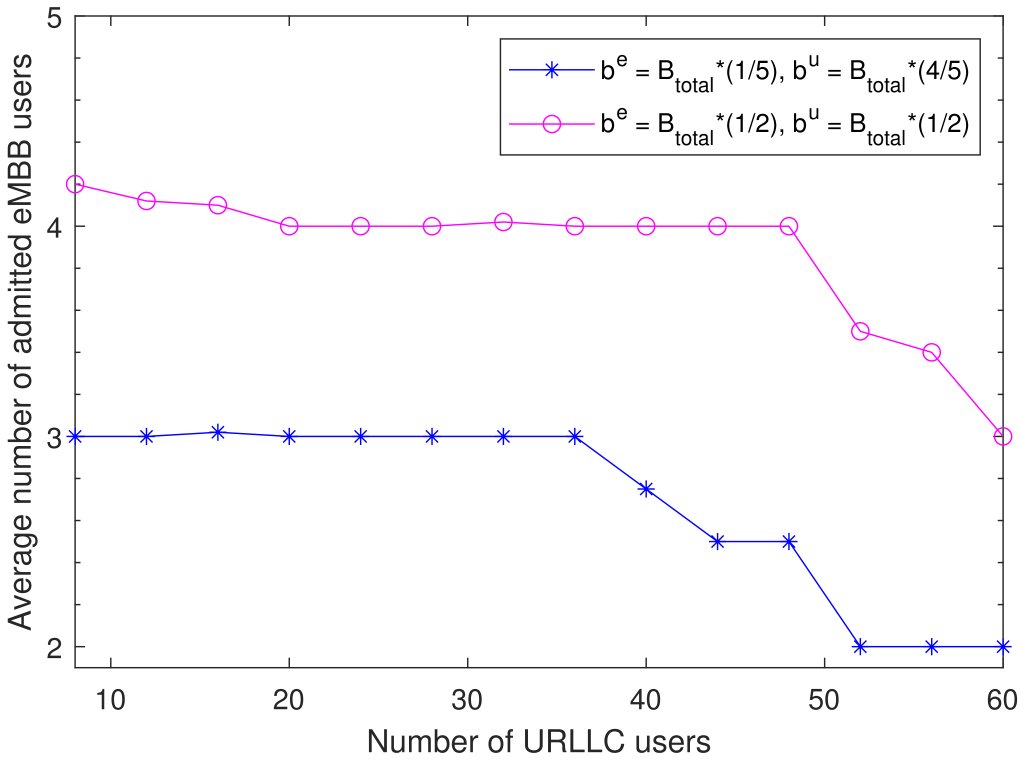

Figure 4.6. Admitted eMBB users versus number of URLLC users in the system.

We can observe from the Figure 4.6, that the number of admitted eMBB users is decreasing with the increase of URLLC users in the system. Admitted number of eMBB users begin to decrease with the low number of URLLC users with limited bandwidth resources (for the case 1) to eMBB users compared to high bandwidth resource to eMBB users (for the case 2).

From the numerical results, we can conclude that our system can find the possible number of admitted eMBB users with required rate when all URLLC users have satisfied their reliability and latency requirements. Optimal bandwidth allocation between eMBB users and URLLC user will lead to a higher number of admission of eMBB users while satisfying the reliability and latency requirement of URLLC users.

4.3 Limitations during the simulation

We have faced following challenges during the simulations. We have faced difficulties in finding a set of values that make a feasible solution. We have used CVX to solve the admission control problem. CVX is using the successive approximation method to deal with the functions exp, log, log-det, and other functions from the exponential family. This method is less reliable and slower than the methods which are used for other models by CVX [30]. In the constraint (54c) we have an exponential function. Therefore we have faced difficulties during the simulation such as we have faced difficulties of getting results for low bandwidth values.

5 SUMMARY

In this chapter, we present the conclusion of this thesis. Furthermore, we present possible future research directions.

5.1 Conclusion

This research aims to contribute a solution for the admission control problem in 5G networks in which eMBB users and URLLC users coexist. Our target is to maximize the number of admitted eMBB users to the system, who have sufficient data rate, while allocating power, bandwidth and beamforming directions to all URLLC users whose latency and reliability requirements are always guaranteed.

In the first chapter, we have presented the introduction to the thesis, the motivation and the objectives of the thesis. In the second chapter, we have presented the background knowledge and a concise literature survey that is related to this thesis.

In the third chapter, we have presented a solution to our problem. We have considered the downlink of a multi-user MISO 5G network. We have suggested orthogonal spectrum sharing between eMBB and URLLC users to coexist them. The objective of the problem is to maximize the number of admitted eMBB users under four constraints. The constraints are signal-to-interference-plus-noise ratio constraint for eMBB users, signal-to-noise ratio constraints for URLLC users in order to satisfy and high reliability and low latency requirements of URLLC users, transmit power constraint at the base station and bandwidth constraint of the system. We formulated the problem as an ℓ_0 minimization problem. Since it is an NP-hard problem we have used sequential convex programming method to solve the problem. SINR constraint of eMBB user is formulated using the achievable rate of eMBB user which derives through Shannon's capacity. Since URLLC has short packet length we have used the approximation for Shannon's capacity in short blocklength regime to obtain the maximum achievable rate for URLLC users. SNR constraint of URLLC users has formulated using the concept of effective bandwidth to guarantee the latency and reliability requirements of URLLC users.

In the fourth chapter, we have presented the simulation setup and numerical results which prove the correctness and effectiveness of our solution. Numerically we have proved the convergence of our algorithm. With numerical results we have shown that number of admitted users increases with the increase of the total bandwidth of the system and with increase of the bandwidth portion for eMBB users. Furthermore, we have shown that the number of admitted users decreases with the increase of target rate for eMBB users. Moreover, numerical results shows that the number of admitted eMBB users increases with the increase of the maximum power of the base station. Finally, we have proven with the help of numerical results that the number of admitted users is decreasing with the increase of number of URLLC users in the system. From the numerical results, we can conclude that our system can find the possible number of admitted eMBB users with the required rate when the all URLLC users have satisfied their reliability and latency requirements.

5.2 Potential future directions

One possible extension of this research is formulating admission control for multi-cell scenario. Also, we can extend this admission control problem to address with user mobility. Another possible extension is optimal allocation of PRBs for eMBB and URLLC users instead of optimal bandwidth allocation. Further we can extend this problem to allocate separate bandwidth portions orthogonally for each eMBB user. In this thesis, we have considered the admission control problem in 5G networks in which eMBB applications and URLLC applications coexist. We have shown that we cannot admit every eMBB user while satisfying the reliability and latency requirements of the all URLLC users. We have to puncture some number of eMBB users in order to give chance to all URLLC users since the low latency budget of URLLC users. Therefore another possible extension of the research is scheduling the punctured eMBB users in unlicensed spectrum. In other words we can recover the loss rate of eMBB users using unlicensed spectrum. We can extend our research to find the number of eMBB users who are not admitted to the system and schedule them in the unlicensed spectrum.

6 REFERENCES

[1] Osseiran A., Monserrat J.F. & Marsch P. (2016) 5G Mobile and Wireless Communication Technology. Cambridge University Press.

[2] Dahlman E., Parkvall S. & Skold J. (2018) 5G NR The Next Generation Wireless Acess Technology. Academic Press publications.

[3] Alsenwi M., Pandey S.R., Tun Y.K., Kim K.T. & Hong C.S. (2019) A Chance Constrained Based Formulation for Dynamic Multiplexing of eMBB-URLLC Traffics in 5G New Radio. In: 2019 International Conference on Information Networking (ICOIN), pp. 108–113.

[4] Chen-Shang Chang & Thomas J.A. (1995) Effective bandwidth in high-speed digital networks. IEEE Journal on Selected Areas in Communications 13, pp. 1091–1100.

[5] She C., Yang C. & Quek T.Q.S. (2018) Cross-Layer Optimization for Ultra-Reliable and Low-Latency Radio Access Networks. IEEE Transactions on Wireless Communications 17, pp. 127–141.

[6] Esswie A.A. & Pedersen K.I. (2018) Opportunistic Spatial Preemptive Scheduling for URLLC and eMBB Coexistence in Multi-User 5G Networks. IEEE Access 6, pp. 38451–38463.

[7] Karimi A., Pedersen K.I., Mahmood N.H., Pocovi G. & Mogensen P. (2019) Efficient Low Complexity Packet Scheduling Algorithm for Mixed URLLC and eMBB Traffic in 5G. In: 2019 IEEE 89th Vehicular Technology Conference (VTC2019-Spring), pp. 1–6.

[8] Tang J., Shim B., Chang T. & Quek T.Q.S. (2019) Incorporating URLLC and Multicast eMBB in Sliced Cloud Radio Access Network. In: ICC 2019 - 2019 IEEE International Conference on Communications (ICC), pp. 1–7.

[9] Matskani E., Sidiropoulos N.D., Luo Z. & Tassiulas L. (2008) Convex approximation techniques for joint multiuser downlink beamforming and admission control. IEEE Transactions on Wireless Communications 7, pp. 2682–2693.

[10] Gong X., Vorobyov S.A. & Tellambura C. (2011) Joint Bandwidth and Power Allocation With Admission Control in Wireless Multi-User Networks With and Without Relaying. IEEE Transactions on Signal Processing 59, pp. 1801–1813.

[11] Nguyen D.H.N., Bao Le L. & Le-Ngoc T. (2015) Multiuser Admission Control and Beamforming Optimization Algorithms for MISO Heterogeneous Networks. IEEE Access 3, pp. 759–773.

[12] Shashika Manosha K.B., Joshi S.K., Codreanu M., Rajatheva N. & Latva-aho M. (2018) Admission Control Algorithms for QoS-Constrained Multicell MISO Downlink Systems. IEEE Transactions on Wireless Communications 17, pp. 1982–1999.

[13] Tse D. & Viswanath P. (2005), Fundamentals of Wireless Communication.

[14] Cover T.M. & Thomas J.A. (2006) Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing). Wiley-Interscience, New York, NY, USA.

[15] Polyanskiy Y., Poor H.V. & Verdu S. (2010) Channel Coding Rate in the Finite Blocklength Regime. IEEE Transactions on Information Theory 56, pp. 2307–2359.

[16] (accessed 19.9.2019.), Introducing 5G networks – Characteristics and usages (2019). URL: `https://www.gemalto.com/mobile/inspired/5G`.

[17] Kavanagh S. (accessed 19.9.2019.), What is enhanced Mobile Broadband (eMBB). URL: `https://5g.co.uk/guides/what-is-enhanced-mobile-broadband-embb/`.

[18] Popovski P., Trillingsgaard K.F., Simeone O. & Durisi G. (2018) 5G Wireless Network Slicing for eMBB, URLLC, and mMTC: A Communication-Theoretic View. IEEE Access 6, pp. 55765–55779.

[19] Sun C., She C. & Yang C. (2017) Exploiting Multi-User Diversity for Ultra-Reliable and Low-Latency Communications. In: 2017 IEEE Globecom Workshops (GC Wkshps), pp. 1–6.

[20] She C., Yang C. & Quek T.Q.S. (2017) Radio Resource Management for Ultra-Reliable and Low-Latency Communications. IEEE Communications Magazine 55, pp. 72–78.

[21] Sun C., She C., Yang C., Quek T.Q.S., Li Y. & Vucetic B. (2019) Optimizing Resource Allocation in the Short Blocklength Regime for Ultra-Reliable and Low-Latency Communications. IEEE Transactions on Wireless Communications 18, pp. 402–415.

[22] Amjad M. & Musavian L. (2018) Performance Analysis of NOMA for Ultra-Reliable and Low-Latency Communications. 2018 IEEE Globecom Workshops (GC Wkshps) , pp. 1–5.

[23] Chengjian S. & Yang C. (2019), Unsupervised Deep Learning for Ultra-reliable and Low-latency Communications.

[24] F. Kowalewski E.S. (2017) 5G Frame Structure. White Paper .

[25] Pandey S.R., Alsenwi M., Tun Y.K. & Hong C.S. (2019) A Downlink Resource Scheduling Strategy for URLLC Traffic. In: 2019 IEEE International Conference on Big Data and Smart Computing (BigComp), pp. 1–6.

[26] Wai H. & Ma W. (2012) A decentralized method for joint admission control and beamforming in coordinated multicell downlink. In: 2012 Conference Record of the Forty Sixth Asilomar Conference on Signals, Systems and Computers (ASILOMAR), pp. 559–563.

[27] Boyd S. (accessed 14.06.2019.), Sequential Convex Programming. URL: `https://stanford.edu/class/ee364b/lectures/seq_slides.pdf`.

[28] Lipp T. & Boyd S. (2016) Variations and extension of the convex-concave procedure. Optim. Eng. 17, pp. 263–287.

[29] Yang W., Durisi G., Koch T. & Polyanskiy Y. (2014) Quasi-Static Multiple-Antenna Fading Channels at Finite Blocklength. IEEE Transactions on Information Theory 60, pp. 4232–4265.

[30] (accessed 19.8.2019.), CVX Users' Guide – Advanced topics. URL: `http://cvxr.com/cvx/doc/advanced.html`.