Enabling Optimized Resource Management in 5G MAC Layer to achieve Lower URLLC Latency and Enhancing EMBB Data Rate

by

Ali FEIZ

THESIS PRESENTED TO ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
IN PARTIAL FULFILLMENT OF A MASTER'S DEGREE
WITH THESIS IN ELECTRICAL ENGINEERING
M.A.Sc.

MONTREAL, NOVEMBER 28, 2022

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC

BOARD OF EXAMINERS

THIS THESIS HAS BEEN EVALUATED

BY THE FOLLOWING BOARD OF EXAMINERS

Mr. Zbigniew Dziong, Thesis supervisor
Department of Electrical Engineering, École de technologie supérieure

Mr. Jean-Charles Grégoire, Co-supervisor
Department of Electrical Engineering, INRS

Mr. Kim Khoa Nguyen, President of the board of examiners
Department of Electrical Engineering, École de technologie supérieure

Mr. Michel Kadoch, Member of the jury
Department of Electrical Engineering, École de technologie supérieure

THIS THESIS WAS PRESENTED AND DEFENDED

IN THE PRESENCE OF A BOARD OF EXAMINERS AND THE PUBLIC

ON "DEFENSE DATE"

AT ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisor, Prof. Zbigniew Dziong, for enduring encouragement during my Master's thesis and related research, along with his persistence, inspiration, and profound expertise. His encouragement was precious within that research project and the composition of this thesis.

Aside from my director, I gratefully acknowledge Prof. Jean-Charles Grégoire, my co-supervisor, for all your efforts in assisting me. His informative comments, motivation, and questions prompted me to broaden my study into new areas. Your approach was brilliant, enabling me to study in a manner I had never noticed before.

In all of my endeavors, my family and friends have always supported and encouraged me. Your belief in me and desire for the best are always evident to me. Thanks to my parents and brother, I have learned how to be happy, to be self-aware, and to appreciate others first and foremost.

Permettre l'ordonnancement des ressources radio dans la couche MAC de la 5G pour réduire la latence

Ali FEIZ

RÉSUMÉ

La 5G devrait permettre la mise en œuvre des trois services suivants : communications ultra-fiables à faible latence (uRLLC), haut débit mobile amélioré (eMBB) et communications massives de type machine (mMTC). uRLLC est essentiel à la prise en charge des dispositifs IoT dans les domaines du traitement, de la distribution d'électricité et des services médicaux. eMBB est une étape importante par rapport aux réseaux 4G actuels qui offre des débits de données plus élevés pour les systèmes de réseaux cellulaires 5G. La qualité de service (QoS) et la qualité d'expérience (QoE) indiquent la performance globale des réseaux cellulaires telle qu'elle est perçue par ses utilisateurs. Pour atteindre les objectifs de la 5G, la qualité de service et la qualité d'expérience doivent être ajustées pour chacun des services de la 5G.

Cette étude se concentre sur la différenciation QoS/QoE pour améliorer T_{Radio} qui représente le délai de transmission des paquets entre le gNB et l'UE. Cette amélioration est obtenue en proposant une technique d'ordonnancement qui permet de réduire la latence uRLLC tout en offrant un débit de données plus élevé pour les applications eMBB. Cet ordonnancement, appelé Enhanced Delay-based QoS Aware Scheduling (EDQAS), est un nouvel algorithme d'ordonnancement MAC de liaison descendante qui détermine quels blocs de ressources physiques sont alloués à chaque paquet.

Le modèle EDQAS incorpore trois nouvelles caractéristiques. La première consiste en un mécanisme de contrôle efficace des délais (EDC) pour reconnaître le type de trafic. La deuxième met en œuvre une méthode de perforation des ressources inspirée de Knapsack pour servir plus rapidement les paquets uRLLC. La troisième met en œuvre la méthode LDI (least delay increase) en tant que processus de coupe pour rejeter certains paquets qui ne peuvent pas remplir les critères de retard et d'overhead.

Les résultats de la simulation, axés sur les services uRLLC et eMBB, montrent que l'algorithme EDQAS améliore considérablement le délai, le bon débit et le taux de perte de paquets des services.

Mots-clés: uRLLC, 5G, ordonnancement, qos, liaison descendante, latence, mac

Enabling Optimized Resource Management in 5G MAC Layer to achieve Lower URLLC Latency and Enhancing EMBB Data Rate

Ali FEIZ

ABSTRACT

5G is expected to enable the following three services; ultra-reliable low latency communications (uRLLC), enhanced mobile broadband (eMBB), and massive machine-type communications (mMTC). uRLLC is essential to the support of IoT devices in processing, electricity distribution, and medical services. eMBB is a significant milestone from current 4G networks that provides higher data speeds for 5G cellular network systems. Quality of service (QoS) and Quality of Experience (QoE) indicate the overall performance of cellular networks as perceived by its users. To achieve the 5G goals, the QoS and/QoE need to be adjusted for each of the 5G services.

This study is focused on QoS/QoE differentiation to improve T_{Radio} that represents the packet transmission delay between the gNB and UE. This improvement is achieved by proposing a scheduling technique that provides lower uRLLC latency while offering higher data rate for eMBB applications. This scheduling, named Enhanced Delay-based QoS Aware Scheduling (EDQAS), is a new downlink MAC scheduling algorithm that determines which physical resource blocks are allocated to each packet.

The EDQAS model incorporates three novel features. The first consists of an efficient delay control (EDC) mechanism to recognize the traffic type. The second implements a resource puncturing method inspired by Knapsack to serve uRLLC packets faster. The third implements the least delay increase (LDI) method as a cutting process for dismissing certain packets that cannot fulfill the delay and overhead criteria.

The simulation results, focused on uRLLC and eMBB services, show that the EDQAS algorithm significantly improves the delay, goodput, and packet loss ratio of the services.

Keywords: uRLLC, 5G, scheduling, qos, downlink, latency, mac

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

LIST OF ALGORITHMS

Page

LIST OF ABBREVIATIONS

3GPP	3rd Generation Partnership Project
ARQ	Automatic Repeat Request
BLER	Block Error Rate
CDF	Cumulative Distribution Function
CQI	Channel quality indicator
DQAS	Delay based QoS Aware Scheduling
DRB	Data Radio Bearer
EDQAS	Enhanced delay based QoS aware scheduling
EXP-PF	Exponential Proportional Fairness
EDC	Efficient delay control
eMBB	enhanced Mobile Broad Band
FDD	Frequency-Division-Duplex
FEC	Forward error correction
gNB	Next Generation NodeB
HARQ	Hybrid Automatic Repeat Request
ITU	International Telecommunication Union
IoT	Internet of Things
LDI	Least delay increase
MAC	Medium Access Control

MLWDF	Modified Largest Weighted Delay First
MCS	Modulation coding scheme
mMTC	massive Machine Type Communication
MME	mobility management management entity
NR	5G New Radio
NRT	Non Real Time
PDCP	Packet data convergence protocol
PDSCH	Physical downlink shared channel
PF	Proportional Fairness
PRB	Physical Resource Block
QoS	Quality of service
RT	Real Time
RLC	Radio Link Controller
SGSN	Serving GPRS support node
SNR	Signal-to-Noise-Ratio
TDD	Time-Division-Duplex
TTI	Time Transmission Interval
UE	User equipment
uRLLC	ultra Reliable Low Latency Communication

LIST OF SYMBOLS AND UNITS OF MEASUREMENTS

α	Delay-throughput optimum
$D_{HOL,j}$	Packet j head of line delay (buffer delay)
$\overline{D_{HOL}}$	Average HOL delay
D_{inc}	Increase on buffer delay
$D_{uRLLC,j}$	delay ∀ j ∈ uRLLC
$D_{eMBB,j}$	delay ∀ j ∈ eMBB
D_{size}	Payload size of uRLLC traffic
δ	incline ratio
E	eMBB UE at TTI t
γ	$D_{uRLLC,j}$ equilibrium factor
$I(min(D_{inc})$	Delay increase index
$I(min(Thr_{dec})$	Throughput decrease index
$I(min(Dr_{inc})$	Drop increase index
i	Punctured resource index for last user
M_{ji}	Priority metric of packet j on RB i
$maxM_{ji}$	Max M[j][i] in TTI
N_{TTI}	Number of TTIs
O_{RB}	Selected RB for puncturing
P	Recording punctured resources from each eMBB UE

R	Data rates at TTI for eMBB UEs
R_a	The eMBB data rate and SNR of uRLLC
$r_U(k)$	Data rate of uRLLC user U at resource block k
SNR_E	SNR values of eMBB UE
SNR_U	SNR values of uRLLC UE
σ_j	Stabilized delay-drop index
Thr_{dec}	Decrease on average throughput
U	uRLLC traffic demanding immediate service
μ_j	Packet j data rate
$\overline{\mu}$	Average data rate

INTRODUCTION

5G is expected to enable the following three services; ultra-reliable low latency communications (uRLLC), enhanced mobile broadband (eMBB), and massive machine-type communications (mMTC). Delivering these services with required QoS/QoE characteristics is a critical objective of 5G cellular networks. Real-time communication, telesurgery, industrial automation, and video conferencing demand high throughput, dependability, and lower latency in a 5G network. However, providing these requirements might be limited because they interfere with each other and it is mainly the MAC layer that can address these issues via improved scheduling (Parvez *et al.* (2018)).

There are several strategies presented in the literature to overcome the MAC layer issues. Moreover, several downlink scheduling methods, resource management, and routing techniques are addressed. Nevertheless, algorithms for dynamic resource allocation in MAC layer are still not standardized since each manufacturer uses different approaches. Therefore, challenges in developing and evaluating different MAC algorithms continue. In particular, current resource management algorithms in the MAC layer do not allow fast and reliable transmission for many upcoming applications. In this study we propose the Enhanced Delay-Based QoS Aware Scheduling (EDQAS) scheduler that is designed to meet the needs of uRLLC applications that require faster and more reliable transmission methods and enhance the eMBB data rate.

Multiple studies are applied to determine 5G scheduler effectiveness for different channel conditions, networks, and parameter specifications. Several emerging innovations in 5G radio schedulers and modern QoS systems are designed to operate the specialized medium access control layer and illustrate the potential of improved high-layer scheduling capability. The proposed EDQAS algorithm operates inside the RRH, as illustrated in Figure 0.1, to differentiate QoS of uRLLC, eMBB, and mMTC services and prioritizes uRLLC services to serve it faster.

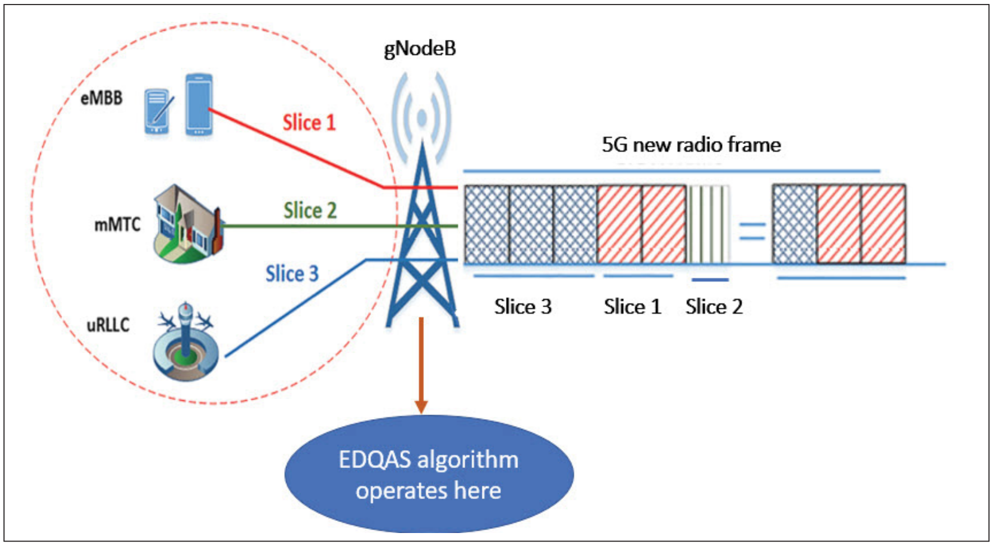

Figure 0.1 Resource management in 5G new radio
Taken from AlQahtani *et al.* (2020)

MAC Scheduler

Data is moved between MAC and physical layers through transport blocks. The transport block consists of a physical resource block(a block of subcarriers) and modulation coding scheme that are further combined by a cyclic redundancy check. Before being mapped onto the physical downlink shared channel for transfer via the air interface, a transport block goes through the physical layer processing at the transmitter. The scheduler aims to satisfy the QoS goals for each data radio bearer that is associated with a particular service type. Each user has at least one default data radio bearer when a new end-to-end packet session is established in the 5G RAN. By assigning distinct radio bearers to packets with different QoS criteria, radio bearers enable the differentiation of packets with different QoS demands (Pedersen et al., 2018).

Problem statement

Below is a list of the challenges associated with resource blocks management of the bandwidth to uRLLC, eMBB, and mMTC aligned with the goals of 5G:

Challenge 1: The first challenge is to determine appropriate KPIs that can characterize well the performance of MAC scheduler since the previously proposed KPI sets are not very good for the addressed problem. This is important for accurate and monolithic resource management.

Challenge 2: Another challenge is to design fair resource allocation policies for different 5G slices(services) that have different radio access technology parameters and predefined quality requirements. However, fulfilling all performance requirements for uRLLC, eMBB, and mMTC at the same time is very difficult. Additionally, 5G slices require specialized traffic management to minimize network congestion.

Objectives

This research aims to reduce uRLLC latency and improve eMBB data rates in a 5G network through dynamic scheduling of resource allocation for uRLLC, eMBB, and mMTC applications. Furthermore, the proposed scheduler strives to achieve the QoS objectives for all services and related data radio carriers utilized by UEs. Therefore, each UE requires a data radio bearer (DRB) to establish a new end-to-end (E2E) packet connection in the 5G RAN domain. For differentiation of QoS for services with respect to latency, packet drop ratio, and throughput criteria, different DRBs should be configured.

Methodology

This project proposes the Enhanced Delay-based QoS Aware Scheduling or **EDQAS** algorithm to find an efficient resource scheduling scheme for 5G uRLLCs. The EDQAS aims to minimize latency of uRLLC traffic while maintaining an acceptable quality of service for the eMBB data rate. An attractive option for RAN solutions is modifying the physical air interface. A new MAC scheduler for the downlink 5G channel is proposed in EDQAS. First the efficient delay control (EDC) mechanism is developed to meet the uRLLC delay requirements.

Then, resource blocks (RB) are optimally selected during the puncturing phase based on the channel conditions. The smallest volume of bandwidth that can be allocated to a user is known as a RB. uRLLC channel conditions are measured using SNR and tagged as the weight. Higher SNR means a higher weight that reflects the impact of puncturing this RB on eMBB UEs' goodput. A uRLLC service uses the Knapsack decision parameter to determine the maximum payload size to be transmitted. The gNodeB punctures previously allocated eMBB resources to serve uRLLC traffic faster.

Further, the least delay increase (LDI) mechanism is designed to schedule services and improve delay, goodput, and drop ratio. The LDI mechanism allows EDQAS to schedule services, providing a steady and reasonable balance between latency, packet loss ratio, and goodput for both uRLLC and eMBB traffic.

Results

To evaluate the new EDQAS method, we compared the EDQAS algorithm performance metrics (delay, goodput, and packet loss ratio) with three selected benchmark quality of service-aware scheduling schemes. The performance evaluation focused on the effectiveness of EDQAS in minimizing latency. Another investigated aspect is the capability of maintaining QoS levels for eMBB data rate under high network loads.

In the tested scenarios with high number of UEs, the EDQAS algorithm improved the uRLLC delay in the range of 2-6% when compared to other algorithms. The results also show that EDQAS improved the eMBB delay in the range of 2-3% for medium traffic usage.

The EDQAS has also improved the uRLLC goodput by 2-6% compared to the benchmark algorithms for high network loads. Moreover, for a large number of users, EDQAS significantly improved eMBB goodput in the range of 6-37%.

Regarding the packet losses, EDQAS reduced the uRLLC packet loss ratio by 6-15% for scenarios with a large number of users when compared to the benchmark scheduling algorithms. For eMBB service, the packet loss ratio was reduced by 3-11% for scenarios with many users.

Thesis Outline

There are three sections in this research: Section 1 provides a summary of 5G characteristics and research areas to be explored for achieving lower latency in 5G uRLLC. In section 2, we describe in detail the proposed EDQAS scheduler. The new 5G air simulator is discussed in Section 3. The remaining part of section 3 explores new features designed for testing 5G systems. Section 4 presents numerical results and performance analysis. Conclusion and future research are discussed in section 5. Appendix 1 explains the implementation of the 5G air simulator.

CHAPTER 1

LITERATURE REVIEW

1.1 Review of 5G characteristics

5G wireless technology is meant to revolutionize telecommunications. This technology will enable faster speeds, lower latency, and many other advantages that will facilitate future technologies. 5G operators are predicting theoretical speeds of up to 10Gbps, which is nearly 100 times faster than 4G speeds of 100Mbps.

The 5G network will transform wireless communications. Unlike previous wireless technologies, which focused solely on mobile broadband, 5G offers a broader range of capabilities. Fast 5G networks can provide a dependable, real-time connection to IoT devices, D2D communications, remote medical care, automated public transportation, augmented reality, and other technologies that rely on real-time control and dependable connectivity. With mobile phones and tablets, users benefit from faster download speeds and lower power consumption. As a new evolution of wireless communication technologies, 5G NR is intended for both mobile users and commercial applications. 5G has introduced new use cases such as uRLLC, eMBB, and mMTC. The demands for these services are higher than the actual latency of 1ms, and the bit rate is higher at 1Gbps compared to LTE (Mataj, 2020).

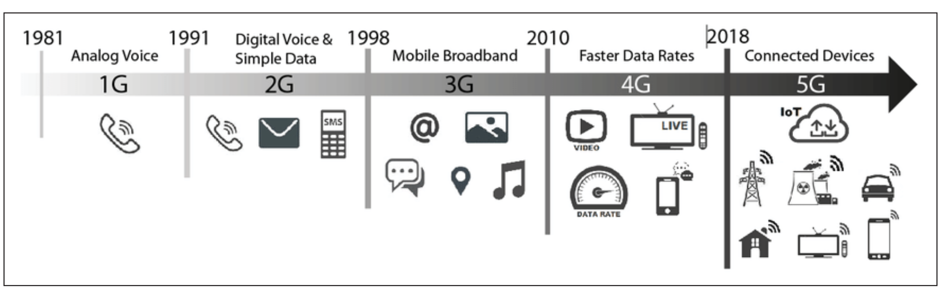

Figure 1.1 1G to 5G network evolution
Taken from Guevara *et al.* (2020)

1.1.1 Ultra-reliable Low-latency Communications (uRLLC)

These applications are designed explicitly for latency-sensitive devices that need maximum reliability, such as industrial automation, autonomous driving, remote surgery, augmented reality, virtual reality, and tactile interface as emerging uRLLC applications (Li et al., 2018).

1.1.2 enhanced Mobile Broadband (eMBB)

eMBB is a high-bandwidth service appropriate for web searching and video streaming. This feature is an evolution of LTE advanced technology designed to increase data rates offered to mobile customers to alleviate network congestion, especially in densely populated areas. eMBB offers improved connectivity, capacity, and user mobility in 5G networks (Zaidi et al., 2018).

1.1.3 massive Machine Type Communications (mMTC)

This service is characterized as narrowband internet connectivity, often provided by IoT devices that connect many devices to detect, measure, and monitor innovative grid services. The traffic type is inconsistent within a particular time. Apart from the vast connection, it is worth noting that mMTC devices deal with battery saving and tend to run for a long time, requiring comprehensive coverage and extensive indoor penetration (Osseiran et al., 2016).

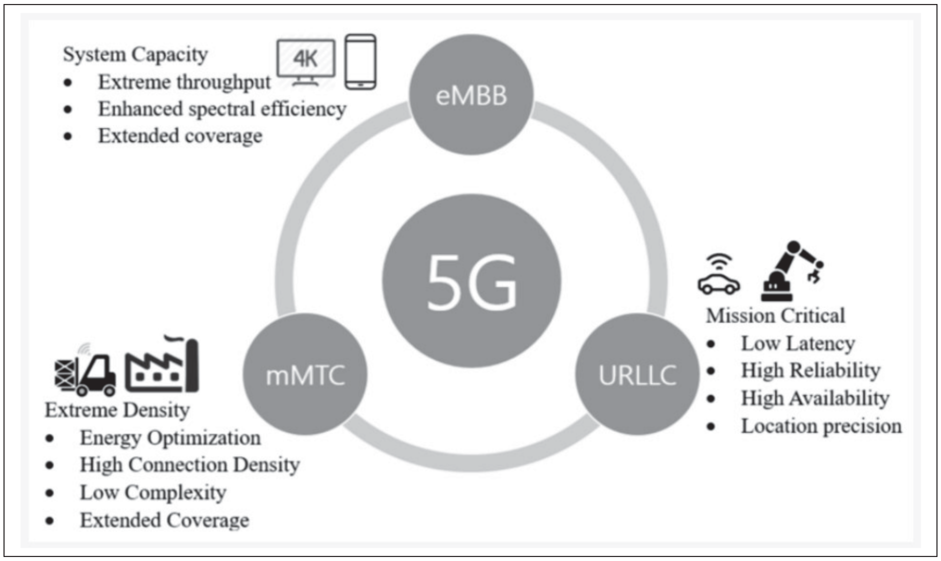

Figure 1.2 5G distinct characteristics
Taken from Siddiqi *et al.* (2019)

1.2 5G low latency services

Many industries, such as robotics, transportation, and medicine, are susceptible to latency. Moreover, The internet of things is rapidly becoming a reality, allowing everything to be connected from anywhere in the world. Global connections of watches, glasses, smart homes, detectors, self-driving cars, and robots enhance our quality of life ((Schulz,Riedel et al., 2017),(Sarwat et al., 2018)). However, the resource scarcity and time-dependent characteristics of uRLLC complicate its handling of arrival services at the gNB. According to the 5G features definitions, the uRLLC services should be provided by the gNB within a Transmission Time Interval (TTI). Retransmission of data will be much faster with 5G NR since HARQ is much more efficient than LTE. HARQ is an acronym that is composed of automatic repeat requests (ARQ) and forward error correction (FEC). If the sender fails to receive an acknowledgment before the timeout, the receiver discards the erroneous packet, prompting the sender to retransmit it. The FEC method relies on the transmitter sending redundant data to the receiver, which recognizes only those portions of it that do not contain errors. 5G NR deployment of HARQ enhances uRLLC and eMBB. 5G NR supports retransmission at three layers: MAC, RLC, and PDCP.

The HARQ retransmissions accelerate MAC layer operations and improve low-latency applications' performance. Providing high speeds as high as 1Gbps for 5G applications requires highly stable connections to prevent TCP congestion avoidance. HARQ quick retransmissions can be combined with RLC-guaranteed packet delivery to achieve this objective. The performance can be improved, but additional feedback signaling raises mMTC delays((Rao et al., 2018),(Dahlman et al., 2020)).

3GPP offered two scheduling algorithms for managing uRLLC traffic. A reservation-based scheduling strategy is known as reservation-based scheduling, while a fast or preemptive scheduling strategy is called fast scheduling. For resource management of unexpected traffics in the first method, a uRLLC reservation-based frame is assigned by gNB. It can apply whether in static or dynamic resource allocation. The static allocation technique sends the frame structure that holds the transmission settings in a periodic form. Unlike static reservation, dynamic reservation communicates the frame structure continuously to the UE. This strategy generates control signaling overhead, and in the absence of incoming uRLLC data, the allocated resources for the uRLLC services will be lost. The EDQAS utilizes the second method, called preemptive scheduling, as the new proposed algorithm in this research to achieve improvement in faster serving of incoming uRLLC traffic through mini-slot scheduling based on short TTIs of 2,4, and 7 OFDM symbols. While this strategy might affect ongoing transmissions of other services such as eMBB and mMTC, it is perceived as a better solution to fulfill the uRLLC goals.

1.2.1 5G services

Here in we introduce a number of services which could benefit from 5G's new traffic classes.

Smart Transportation

Lower latency communication is necessary for the optimization of road traffic. Coordination among automated vehicles is essential for operations like dedicated lanes and racing. uRLLC determines a delay threshold of up to 10 ms for data transmissions of self-driving vehicles. Applications like augmented reality and see-through vehicles are also allowed to stream video with a maximum delay of 50 milliseconds (Parvez *et al.*, 2018).

Figure 1.3 Smart Transportation
Taken from Siddiqi *et al.* (2019)

Teleconferencing and Robotic systems

Using remote-controlled robots in hazardous environments is an emerging industry. Drones and video conferencing technologies require remote control and real-time synchronized visual support. It is reasonable to expect a network response time of a few milliseconds ((Fettweis et al., 2014), (Simsek et al., 2016), and (Schulz *et al.*, 2017)).

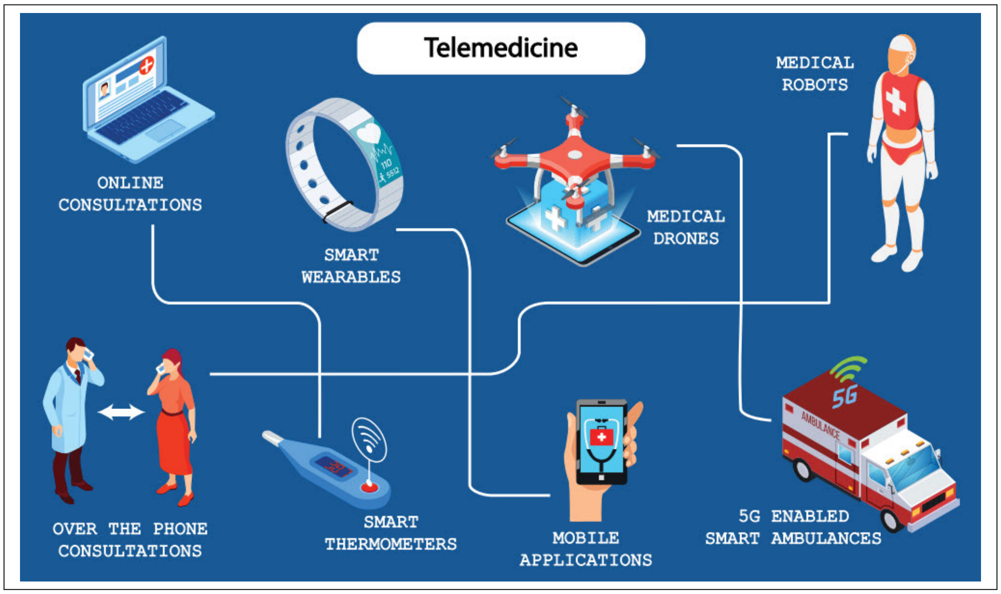

Figure 1.4 Telemedicine and Robotic systems
Taken from Chamola *et al.* (2020)

Virtual Reality

When manipulating objects with specific technologies such as telesurgery, it is essential to maintain extremely high accuracy levels. Virtual reality enables several users to interact through physical VR simulations in physical VR spaces as part of a shared haptic experience. Networked connectivity needs to be improved in terms of low latency consistency and user interaction with streamlined workflows ((Fettweis *et al.*, 2014),(Simsek *et al.*, 2016), and (Kasgari et al., 2019)).

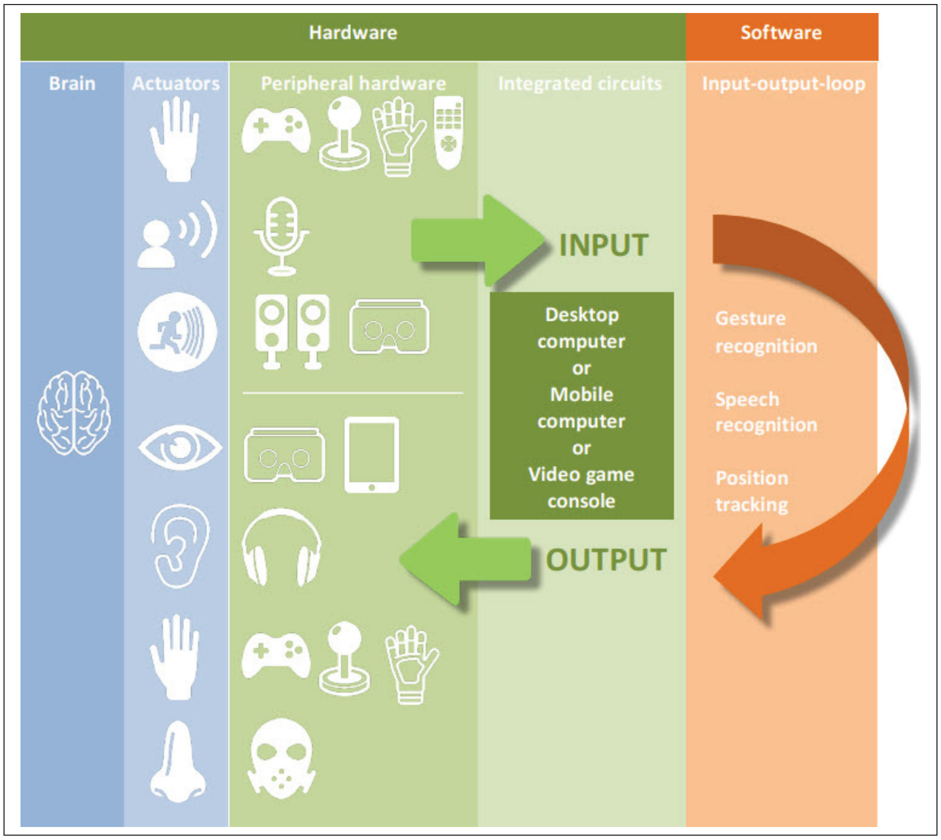

Figure 1.5   VR functional principle

Gaming

Professional gaming has a purpose other than amusement. Games like these are designed to overcome obstacles and motivate players to achieve goals that can be applied in various environments, including education, training, simulation, and health care. Almost over 30-50 ms of network latency inevitably lead to a drastic reduction in game quality and gaming experience rankings. For tangible social interaction with rising visualization, a 1 ms round trip time is proposed(Simsek *et al.*, 2016).

Smart Grid

Reliability and latency criteria for the smart grid are very demanding. Using dynamic control, we can turn on and off providers in 100 ms by setting a delay in the E2E direction. For continuous power supplier co-phasing (i.e., generators) and dynamic activation and deactivatiExampleE2E

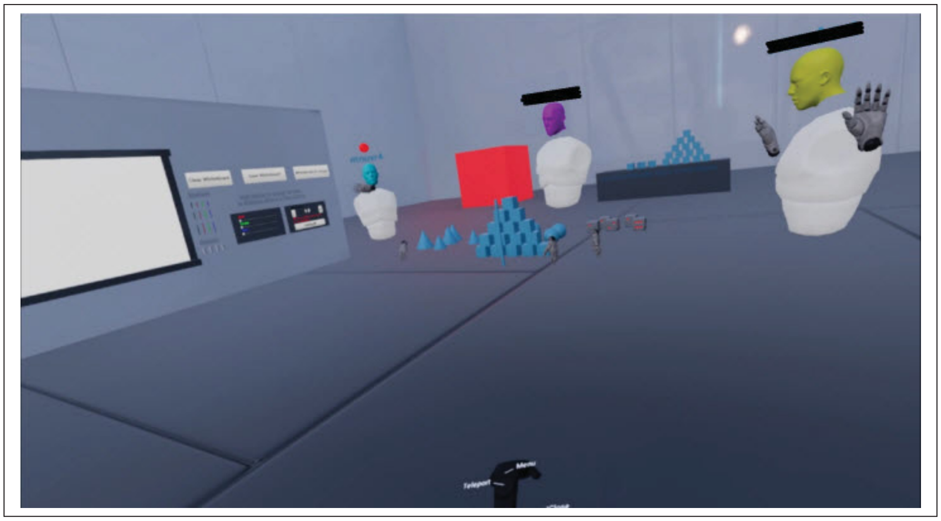

Figure 1.6 VR operation connection
Taken from Huseb *et al.* (2018)

Figure 1.7 Legends World Championship
Taken from Rivas *et al.* (2018)

delay of 1 ms is required. Data speed in the range of 1500 kbps is required to achieve wide-area spatial awareness ((Meng et al., 2019) and (Hovila et al., 2019)).

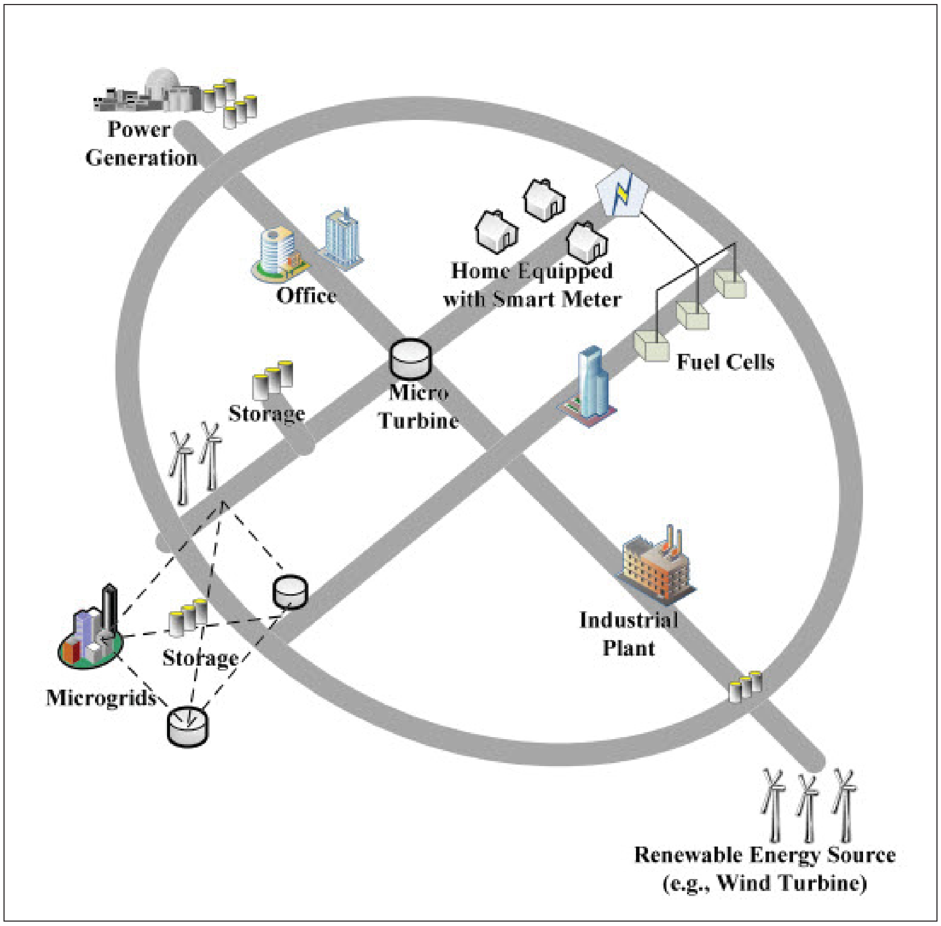

Figure 1.8 Smart grid characteristics
Taken from Kayastha *et al.* (2014)

1.3 Cellular network latency

URLLC latency measures how long it takes to transfer packets from serving to the user equipment. The latency domain of a cellular network includes user plane and control plane latency. This research concentrates on 5G U-Plane latency and packets' IP layer delivery time.

An idle device indicates that the radio resource control (RRC) is not connected. After establishing an RRC connection, the UE switches to connected mode. The user plane is a primary target for low-latency communication due to its significant influence on latency (Garcia-Perez et al., 2016).

Cellular network packet transmission latency can be affected by RAN, backhaul, and core networks. The entire one-way transmission may be stated below for a cellular network.

T = $T_{Radio} + T_{Backhaul} + T_{Core} + T_{Transport}$

Where

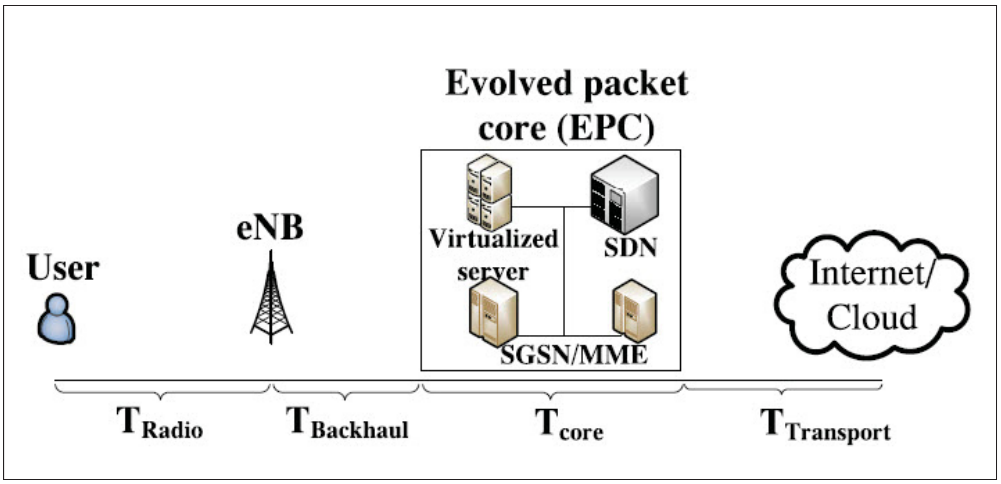

Figure 1.9 E2E latency of packet transmission
Taken from Parvez *et al.* (2018)

T_{Radio} represents the packet transmission delay between a gNB and a UE induced by a physical layer connection and measured by a gNB. This calculation includes processing time at the gNB/UE, re-transmissions, and propagation delay. Processing delays at the eNB can be attributed to channel coding, scrambling, cyclic redundancy checks (CRCs), and OFDM signal synthesis ((Agyapong et al., 2014), and (Pocovi et al., 2016)).

$T_{Backhaul}$ connects the gNB to the Evolved Packet Core (EPC) network. Copper, microwave, or optical fibers are usually used to link the core network to the gNB. Microwaves, in particular, produce lower latency than optical fibers since air has a refractive index of 1.0003, while fiber optic cable has a refractive index of about 1.5. A material with a higher refractive index has a slower speed of light. Therefore, transmission over microwaves is much faster than over fiber optics (Wilner et al., 1976).

T_{Core} is the computation time required by the core network. The service is established with the help of key network elements, including MME, SGSN, and SDN/NFV.

$T_{Transport}$ refers to the time measurement of data from the core network to the service provider. Distance, bandwidth, and communication protocol are frequently the causes of delays. Since this research focused on T_{Radio}, I will elaborate this further. T_{Radio} comprises following components:

$$T_{Radio} = T_{ttt} + T_{prop} + T_{proc} + t_{rtx} + T_{sig} \tag{1.1}$$

- T_{ttt} represents the transmission time for latency.
- T_{prop} indicates the transmitter-to-receiver signal propagation time.
- T_{proc} determines the time for encoding, decoding, and channel estimation during the first transmission.
- T_{rtx} is the retransmission time.
- T_{sig} incorporates connection request, scheduling grant, feedback, and queueing delay.

1.4 Achieving low latency in 5G

Improving the physical air interface is a reliable way to improve latency in the RAN domain, both at the MAC (frame) and the physical layer (modulation, channels, . . .).

1.4.1 Frame structure

An LTE radio frame takes 10 ms. This frame has 10 subframes of 1 ms, split into 0.5 ms parts called resource blocks (RB). In one RB, there are 6 or 7 OFDM symbols, and 120 kHz of frequency (12 consecutive 15-kHz subcarriers). Since the sampling interval T_s is $1/F_s$ and each subcarrier spacing f_{OFDM} is 15 kHz, then every T_{OFDM} symbol duration equals to $1/f_{OFDM}$ = 66.67µs, Therein, the FFT size is 2048 and the sampling rate fs equals f_{OFDM} * N_{FFT}= 30.72 MHz. The subcarrier spacing f_{OFDM} can be set to 30 kHz to minimize TTI and achieve lower latency. Thus, $T[OFDM]$ denotes 33.33 s of OFDM, $N[FFT]$ represents 1024, and sampling rate "fs" remains 33.72 MHz, as in LTE. Each frame period of Ts = 10ms is divided into 40 subframes that are 0.25ms long and contain seven symbols each (Guan et al., 2016).

1.4.2 Modulation and coding schemes

Small packet transmission can effectively reduce latency, but it requires effective modulation and coding to achieve optimum reliability. There are three types of encryption techniques suggested for 5G. In small packets, low-density parity check (LDPC) perform more efficiently than turbo codes, while turbo codes are more efficient in large packets (Sybis et al. (2016)). For low latency, small packets are expected, while other factors such as performance and scalability, practical achievement, and adaptability must also be considered (Wang et al. (2017)).

1.4.3 Millimeter wave Communication

Carrier aggregation over the mmWave spectrum is frequently determined as a potential prospective 5G technology promising to deliver tremendous bandwidth and ultra-low latency. The mmWave MAC layer features a different frame architecture with excellent features such as flexible and reduced transmission intervals, dynamic subcarrier spacing, and directional multiplexings. It outperforms the restrictions of ultra-low latency associated with many subscribers and transmission time intervals (Dutta et al. (2019)).

An adaptable MAC layer and low latency core network architecture should all be considered in MAC layer design. Short symbol periods and low latency mmWave MAC are a few approaches to improve these essential design areas(Ford et al., 2017;Sutton et al., 2019).

1.4.4 Massive MIMO

MIMO (Multiple-Input-Multiple-Output) has played an essential role in developing wireless technology, both theoretically and practically. 4G and 5G networks are predicted to be widely adopted by MIMO technologies. For instance, LTE-Advanced enables spatial multiplexing for both FDD and TDD. MIMO, known as "Stronger Antenna Systems" and "Full-Dimension MIMO" originated in the pioneering work of Marzetta and represented a fresh approach to the existing process by utilizing a sufficient number of antennas (Marzetta (2010)).

MIMO becomes quite interesting at higher frequencies like millimeter waves (mmWave). Massive MIMO offers a promising solution for higher frequency bands since it allows many adjustable antennas to be fitted into a small space (even at UTs). These massive arrays can provide increased power directivity and reimburse for the problematic load variation at these frequencies (Papadopoulos et al., 2016).

1.4.5 QoS/QoE Differentiation

Decoupling QoS and QoE limitations facilitate lower latency in neurosurgery, 3D virtual world, and 5G applications. mmWave bandwidth and beamforming technology used in 5G will ensure high quality of service and experience without sacrificing resource sharing (Wu et al., 2015). This paper proposes a QoS-aware multimedia scheduling approach for mmWave communications, which maps diversified options to the optimal frequency using propagation examination and collision avoidance techniques. With the help of SDN and cloud technologies, the best quality of service and quality of experience can be achieved. Mobile network operators are exploring video stream routing approaches to improve QoE by reducing jitter, stream bit rate, and initiation delay (Roy et al., 2015).

1.4.5.1 QoS objective per class

High data rates, low latency, and high reliability are all essential for the 5G network to meet user expectations. An adequate amount of capital and operating expenditures associated with network operations is required to reach this level of effectiveness. Figure 1.11 illustrates the evolution from 4G (IMT-advanced) to 5G (IMT-2020) to address the above needs resulting from the rapid growth of data traffic.

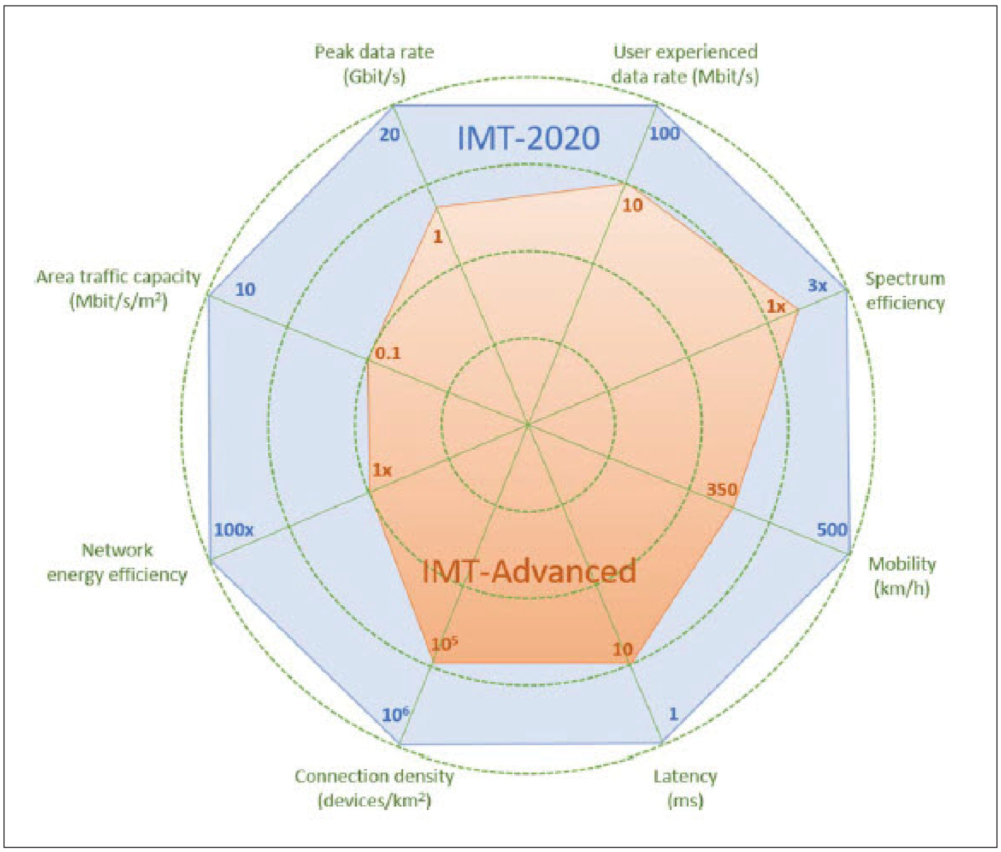

Figure 1.10 IMT-Advanced to IMT-2020
Taken from Navarro-Ortiz *et al.* (2020)

While the ITU categorizes 5G services into different types, addressing these diverse requirements remains challenging (Series, 2015).

ultra Reliable and Low Latency Communication

In industrial automation, power distribution, haptic interfaces, and intelligent transportation, uRLLC is mainly used for time-sensitive applications. In terms of latency, efficiency, and maintainability, uRLLC stands out (Lee & Ko, 2021).

uRLLC's durability level of %99.99 authorizes prompt and accurate data transfer to the highest levels while maintaining integrity, security, and authenticity of small payloads of approximately 256 bits (Pocovi et al., 2018).

enhanced Mobile Broadband

LTE Mobile Broadband technology evolved into eMBB, which provides high-quality video streaming and video games. With eMBB, 5G networks are expected to reach peak data rates exceeding 20 Gbps with only a moderate level of reliability of 10^{-3} packet loss rate. One of the situations where eMBB will be helpful is wireless connectivity and high user density (Series, 2015; Zaidi *et al.*, 2018).

massive Machine Type Communications

mMTC provides connectivity to many devices and is considered a principal IoT integrator. Smart grids, traffic monitoring, and resource management all use mMTC. In addition to massive connectivity, mMTC devices are designed for low battery-power consumption to run for a long time, requiring excellent coverage and extensive interior insertion (Osseiran *et al.*, 2016).

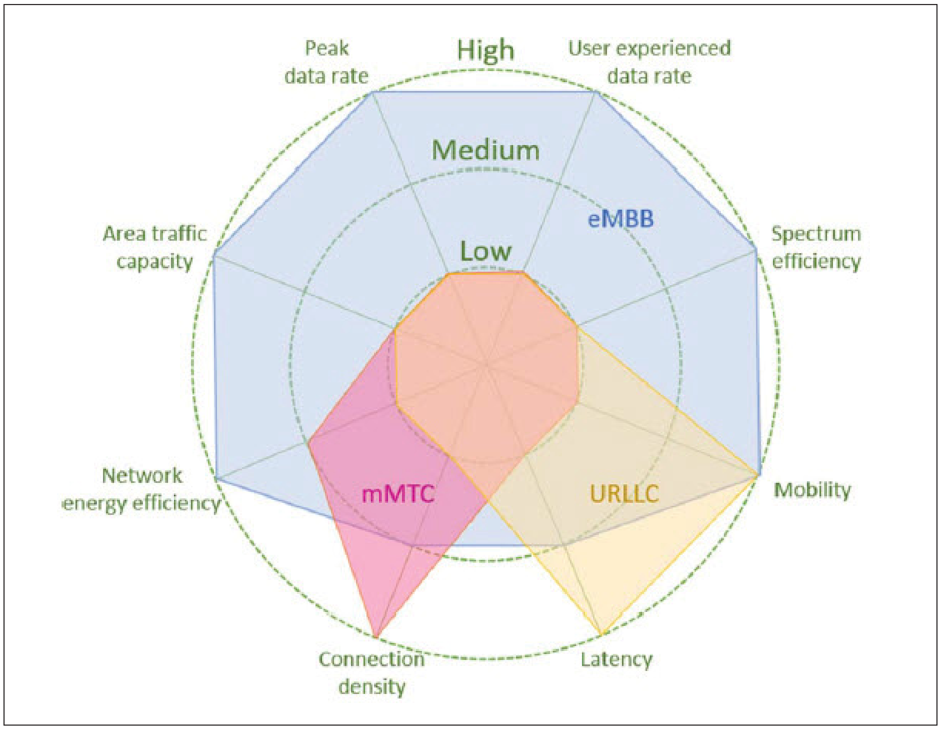

Figure 1.11 5G usage scenarios
Taken from Navarro-Ortiz *et al.* (2020)

CHAPTER 2

PROBLEM STATEMENT AND PROPOSED ALGORITHM

2.1 Enabling dynamic Resource management in the 5G MAC layer

5G technology is typically divided into three subcategories: uRLLC, eMBB, and mMTC. uRLLC requires innovative techniques to meet its stringent latency and reliability requirements. Furthermore, the rapid development of IoT devices has created significant challenges in managing heterogeneous traffic. Radio resource management (RRM) is a critical component of flexibility.

2.2 Problem statement

The main challenge of the uRLLC packet structure is to minimize the transmission time T_{Radio} of a packet. For optimum radio frequency resource utilization in LTE systems, square-shaped packets are used to overcome channel fading. We aim to reach the lower uRLLC latency in this research using a non-square packet architecture along the frequency axis since it reduces transmission delay T_{ttt}. The Knapsack Joint scheduling technique AL-ALI, 2021 is used for minimizing time-to-transmit delays T_{ttt} when both uRLLC and eMBB services are available within one radio resource.

2.3 Problem modeling

As shown in figure 2.1, the number of OFDM signals remains the same while the subcarrier spacing is expanded to minimize the duration of the OFDM symbol. An efficient resource allocation mechanism is achieved by utilizing instant scheduling and numerology type 1 as a frame structure through the new algorithm called EDQAS.

Table 2.1 Simulation parameters for the EDQAS model

Numerology	**class 0**	**class 1**	**class 2**	**class 3**	**class 4**	**class 5**
$Numerology\ scheme(\mu)$	0	1	2	3	4	5
$SCS\ \Delta f = 2^{\mu} \cdot 15[\text{kHz}]$	15	30	60	120	240	480
$RB\ bandwidth[\text{Khz}]$	180	360	720	1440	2880	5760
Symbollength$(10^{-6}s)$	66.67	33.33	16.67	8.33	4.17	2.08
$Time\ slot\ duration\ (ms)$	1	0.5	0.25	0.125	0.0625	0.0313

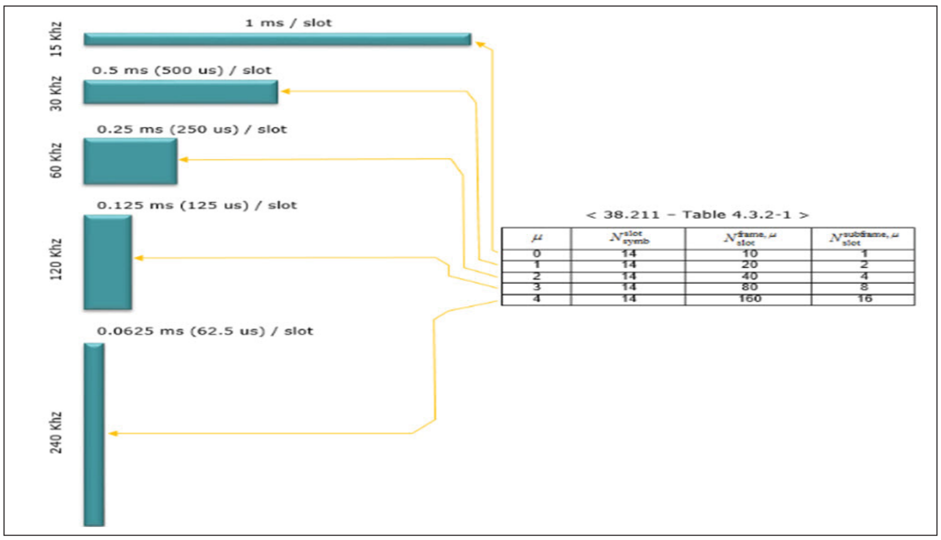

Figure 2.1 5G NR frame structure
Taken from 3GPP (2022)

2.4 System Model

In a 5G network, each UE reports CQI to the evolved gNB over the uplink channel to estimate the link reliability. Therefore, gNB analyzes the reported CQI to recognize the suitable MCS for the downlink transmission to achieve the required BLER for given channel conditions. A noteworthy point to mention is that the gNB is responsible for all the resource allocation and scheduling processes on the system. After gNB receives the CQI response from the relevant UEs, the downlink packet scheduling procedure begins. Figure 2.2 depicts a generic model of EDQAS elements. In EDQAS, the uRLLC should be sent during a time slot already assigned to an eMBB user. The solution describes RB puncturing. RB component values range from 0 to 1, specifying that uRLLC can partially puncture the RB. MAC layer provides an entire scheduling process to update the list of scheduled packets and assign them to a specific bandwidth. EDQAS uses the EDC algorithm as the initial phase to distinguish packet types and classify them with packet size weights (Chagdali et al., 2020; Benjebbour et al., 2018).

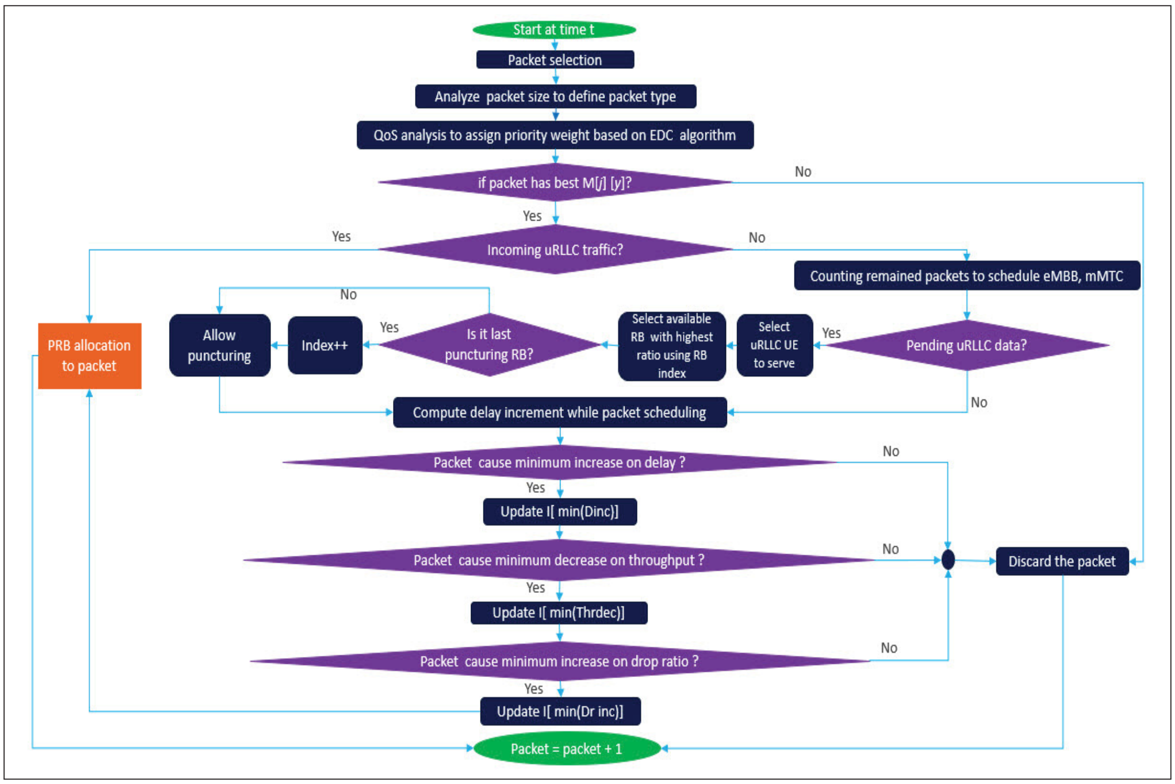

Figure 2.2 Control flow diagram

2.4.1 QoS Analysis in EDC Algorithm

EDQAS comprises the EDC algorithms to differentiate uRLLC, eMBB, and mMTC services and assign a priority weight based on delay. The eMBB and mMTC packets are designed to be served at a lower priority to meet the goals of uRLLC. Algorithm 2.1 shows how this scheduling is achieved step by step. A priority weight is calculated using the buffer delay $D_{HOL,j}$ and assuming j is the uRLLC class. Here we can see the symbols used by EDQAS.

Table 2.2 Control parameters and notations for EDQAS model

Parameters	**Meaning**
α	Delay-throughput optimum
$D_{HOL,j}$	Head of line delay on packet j (buffer delay)
$\overline{D_{HOL}}$	Average HOL delay for available packets
D_{inc}	Increase on buffer delay
$D_{uRLLC,j}$	delay term for packet j, ∀ j ∈ uRLLC
$D_{eMBB,j}$	delay term for packet j, ∀ j ∈ eMBB
D_{size}	Payload size of uRLLC traffic
D_{max}	Max delay bound
δ	Slope coefficient
E	eMBB UE at TTI t
γ	$D_{uRLLC,j}$ balancing factor
$I(min(D_{inc})$	Delay increase index
$I(min(Thr_{dec})$	Throughput decrease index
$I(min(Dr_{inc})$	Drop increase index
i	Punctured resource index for last user
K	list of selected packets to be scheduled at TTI
$M[j][i]$	Priority metric of packet j on RB i
$maxM[j][i]$	maximum value of M[j][i] in TTI
N	List of RBs at TTI
N_{TTI}	Number of TTIs
O_{RB}	Selected RB for puncturing
P	The amount of punctured resources from each eMBB UE
R	data rates for eMBB UEs at TTI t
R_a	SNR ratios for uRLLCs and eMBBs
SNR_e	SNR values of eMBB traffic across all RBs
SNR_u	SNR values of uRLLC traffic over all RBs
σ_j	Delay-drop stabilizing index
Thr_{dec}	Decrease on average throughput
μ_j	Used packet j's data rate
$\overline{\mu}$	A user's average data rate

For a better understanding of the procedure used in this scheduling phase, algorithm 2.1 provides more information. Regardless of increased load and diverse channels, the EDC algorithm is based on delivering real-time service with minimum delay. Based on the buffer delay $D_{HOL,j}$, the priority weight for j corresponds to the uRLLC class:

$$M[j][y] = D_{uRLLC,j}.\mu_j \tag{2.1}$$

$$D_{uRLLC,j} = \frac{\delta}{\exp\left(\alpha \cdot \frac{\overline{D_{HOL}} - D_{HOL,j}}{-\ln\left(D_{HOL,j}\right)\cdot\sqrt{D_{HOL,j}}}\right)} + \gamma \tag{2.2}$$

Note that α is critical in calculating the metric value provided in Eq. 2.8 in such a way to balance goodput (as determined by (μ_j)) and delay ($\overline{DHOL}$ - $DHOL, j$). In this case, σ_j should be carefully chosen to avoid massive packet discarding by not exceeding the D_{max} bound.

$$\alpha = \frac{-\ln\left(\sigma_j\right)}{D_{\max}} \tag{2.3}$$

uRLLC refers to ultralight traffic with fixed packet sizes. This research assumes that UEs always have data to send at each TTI. The logics in Eqs. 2.1, 2.2, and 2.3 are driven by the classic scheduling rule in (Shakkottai et al., 2001). On uRLLC packets, Eq.2.2 imposes a strict delay control on the exponential term $D_{uRLLC,j}$. Then it is recommended to keep the difference between $D_{HOL,j}$ and $\overline{D_{HOL}}$ as low as possible and limiting it by ($\ln\left(D_{HOL,j}\right) \cdot \sqrt{D_{HOL,j}}$).
If D$_{HOL,j}$ is less than $\overline{D_{HOL}}$ by the maximum value of ($-\ln\left(D_{HOL,j}\right) \cdot \sqrt{D_{HOL,j}}$), then packet j is given higher priority as explained in following:

$$\overline{D_{HOL}} - D_{HOL,j} < -\ln\left(D_{HOL,j}\right) \cdot \sqrt{D_{HOL,j}} \tag{2.4}$$

There are several uRLLC services(K > 1) that have similar delay values and are limited by D_{max} at TTI. This circumstance demonstrates the validity of this equation:

$$\overline{D_{HOL}} - D_{HOL,j} < D_{HOL,j} \tag{2.5}$$

It is trivially true that,$[\mathrm{N} < -\ln(N) \cdot \sqrt{N} \forall \mathrm{N} \in [0, 0.1]]$. In this case, the right side of relation 2.4 increases exponentially across all instances $(\mathrm{N}_1, N_2, ..., N_K)$.
Accordingly, relation 6 is:

$$D_{HOL,j} < -\ln\left(D_{HOL,j}\right) \cdot \sqrt{D_{HOL,j}} \tag{2.6}$$

Based on relations 2.5 and 2.6, relation 2.4 applies when 0 < $\mathrm{D}_{HOL,j}$ <D_{max}. The logarithmic factor in Eq.2.2 is associated with δ and γ as adjustable parameters necessary to maintain an appropriate scale for scheduling operations, with values ranging from 0 to 1. Particularly, δ sets the level of the $D_{uRLLC,j}$ service, while γ is a stability coefficient ensuring that $D_{uRLLC,j}$ remains within a given range. By setting δ and γ to [0.5, 0.5], uRLLC packets are adaptable to schedule per delay budgets. In addition, α is denoted as a cutting-sensitive coefficient to optimize the EDC algorithm behavior to preserve minimal latency and drop ratio. Therefore, α manages packet delays. Even though other packets suffer, increasing this parameter reduces delay substantially. Furthermore, during severe traffic congestion, the importance of μ_j increases to guarantee that uRLLC packets are assigned higher weight values for earlier scheduling.
The precedence weight for eMBB UEs in EDC is defined using a dynamic buffer delay bound.

$$M[j][i] = D_{eMBB,j} . \mu_j \tag{2.7}$$

And,

$$D_{eMBB,j} = \alpha \cdot (\overline{D_{HOL}} - D_{HOL,j}) \tag{2.8}$$

During network congestion circumstances, employing $\overline{D_{HOL}}$ as a dynamic threshold to handle large values of $D_{HOL,j}$ to minimize buffer balancing improves burst traffic latency dramatically. When the scheduler adjusts $D_{HOL,j}$ for K packets within a single TTI up to the order of $\overline{D_{HOL}}$, the scheduler prioritizes packets with generally small buffers. It is worth noting that α is the critical key for adjusting either throughput (as determined by (μ_j)) or latency $(\overline{D_{HOL}} - D_{HOL,j})$.

σ_j should be specially designed to prevent massive packet rejection due to time constraints and exceeding D_{max}.

Algorithm 2.1 EDC Algorithm in EDQAS

Initialization;
define N as list of RBs at TTI;
define K as list of selected packets to be scheduled at TTI;
set $M[j][i] \leftarrow 0$, max M[j][i]$\leftarrow$ 0;
set $\overline{D_{HOL}} \leftarrow D_{HOL,j}$;
set $TTI \leftarrow 0$;
for $i \leftarrow 1$ *to* N **do**
    for $j \leftarrow 1$ *to* K **do**
        Compute: α based on Equation (3);
        Update: $\overline{D_{HOL}} \leftarrow \frac{1}{K} \cdot \sum_{j=0}^{K-1} D_{HOL,j}$;
        if $j \in uRLLC$ **then**
            Update $M[j][i]$ according to Equation (2.1);
        else if $j \in eMBB$ **then**
            Update $D[j][i]$ according to Equation (2.8);
        else
            Update $M[j][i] \leftarrow \mu_j$;
        end if
        if *(j not scheduled in TTI && $M[j][i] > maxM[j][i]$)* **then**
            Update $maxM[j][i] \leftarrow$M[j][i];
            Schedule packet j on RB i;
        end if
    end for
end for

2.4.2 Optimised resource puncturing in uRLLC and eMBB services

Puncturing is a dynamic joint scheduling technique that is used to optimize the number of OFDM symbols per subframe and minimize their duration. In other words, slots and mini slots will be allocated dynamically to uRLLC and eMBB services as demonstrated in figure 2.3. The eMBB subscribers are assigned resource blocks at the slot border. The uRLLC has strict latency requirements, so unexpected uRLLC traffic may occur in eMBB time slots already assigned to distinct customers. A subframe consists of 2 slots, each of which contains 14 symbols = 0.5ms. Therefore, the incoming uRLLC traffic is scheduled instantly.

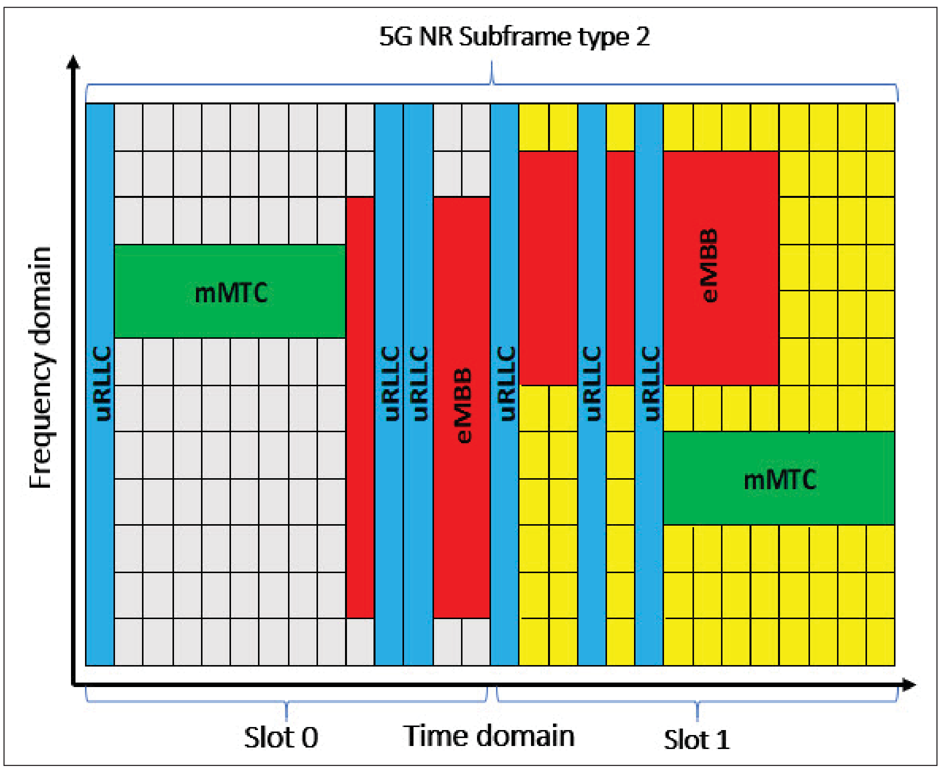

Figure 2.3 uRLLC and eMBB resource puncturing

The eMBB/uRLLC resource allocation problem aims to minimize uRLLC latency while maximizing the eMBB UE data rate. This formulation treats the load as a random variable because uRLLC traffic is stochastic.

The cumulative distribution function can be used to estimate the uRLLC load predictably, provided the uRLLC packet is expressed with a random distribution while ensuring that the planned outage probability does not exceed its formula. This scheduling method minimizes uRLLC traffic latency while maintaining eMBB UE standards and a fair approach to punctured resources across eMBB UEs. Profit and weight objects are associated with each RB in this method. In this example, the weight represents the data rate of the eMBB UE occupying that RB. The profit represents the uRLLC UE's channel conditions at that RB. A better channel condition will result in a greater amount of data will be sent through this RB, which will translate to a greater amount of profit. As a measure of channel condition, the SNR of the selected uRLLC UE is used. Weights reflect the impact puncturing this RB has on the total throughput of eMBB UEs. When solving the problem, the knapsack (AL-ALI, 2021) represents how much payload the user needs to transmit within the current time slot. This part runs after EDC implementation for traffic recognition.

2.4.2.1 Distribution of Resource Blocks to eMBB Users

Maximizing data rate and ensuring user fairness are two objectives of optimizing data rate. All RBs are represented by R = {1, 2, ..., r} and eMBB users are illustrated by E = {1, 2, ..., e}. According to the Shannon capacity model (Alsenwi et al., 2019), an eMBB user's optimum data rate e at time slot t could be estimated as follows:

$$\hat{c}_e(t) = \sum_{r \in \mathcal{R}} f_r x_{e,r}(t) \log_2 \left(1 + \beta \frac{p_e h_e(t)}{N_0 F}\right) \tag{2.9}$$

Here $x_{e,r}$ represents the RB allocation output. $x_{e,r} = 1$ indicates that RB r is assigned to user e and $x_{e,r} = 0$ indicates the reverse situation, f_r indicates the bandwidth allocated to RB r, F represents the overall bandwidth, p_e displays the amount of transmission power by user e, h_e stands for the channel gain of user e, N_0 identifies the noise power and β is refers to bit-error-rate (BER):

$$\beta = \frac{1}{(-\ln(5BER))} \tag{2.10}$$

Here is a description of the average data rate of user e at time t:

$$\bar{C}_e(t) = \omega \bar{C}_e(t-1) + (1-\omega)\hat{c}_e(t) \tag{2.11}$$

And $\omega \in [0, 1]$.

The optimization problem is expressed by the formula:

$$\underset{x}{\text{maximize}} \sum_{e \in \mathcal{E}} \hat{c}_e(t)/[\bar{C}(t)] \tag{2.12}$$

$$\text{Subject to} \sum_{r \in \mathcal{R}} x_{e,r} \leq 1, \forall e \in E \tag{2.13}$$

Condition 2.13 guarantees that each RB is assigned to one user.

2.4.2.2 uRLLC and eMBB transmission planning

The packet loss from eMBB is related to punctured resources at each time slot. The punctured resources defined by $\eta_e(t)$ correspond to eMBB user e's uRLLC traffic. Setting the following parameters determines whether an eMBB user's data rate might be affected during the specified time slot t based on:

$$C_e(t) = (\sum_{r \in \mathcal{R}} f_r x_{e,r}(t) - \eta_e(t)) \log_2 \left(1 + \beta \frac{p_e h_e(t)}{N_0 F}\right) \tag{2.14}$$

Consider that $\psi_e(t)$ represents the proportion of total resources and SNR_e indicates the signal to noise ratio associated with eMBB user e at time slot t, and $\psi_e(t) = \sum_{r \in \mathcal{R}} f_r x_{e,r}(t)\ and\ SNR_e = \beta \frac{p_e h_e(t)}{N_0 F}$.

Then, Eq.2.14 simplifies to:

$$C_e(t) = (\psi_e(t) - \eta_e(t)) \log_2 (1 + SNR_e) \tag{2.15}$$

uRLLC traffic is represented by $[\lambda(t)]$ as a random variable and SNR_u is the signal-to-noise ratio of the punctured mini-slots for uRLLC user e at time slot t. Here is the probability of an outage for the uRLLC:

$$P(E) = P[\sum_{e \in \mathcal{E}} \eta_e(t) \log_2 (1 + SNR_u) < \lambda(t)] \tag{2.16}$$

As mentioned earlier, the uRLLC/eMBB scheduler focuses on improving the data rate of eMBB and optimizing the transmission delay of uRLLC. In addition, it may be expressed as follows:

$$\underset{\eta}{\text{maximize}} \sum_{e \in \mathcal{E}} C_e(t) \tag{2.17}$$

$$\text{Subject to } P(E) \leq \tau \tag{2.18}$$

Whereas,

$$\eta_e(t) \leq \psi_e(t), \forall e \in E \tag{2.19}$$

Where τ is the maximum acceptable probability of failure, Condition 2.18 symbolizes the uRLLC traffic predictability(i.e., 10ms), and limitation 2.19 guarantees the resources assigned to uRLLC traffic are fewer than the available resources.

2.4.2.3 Keeping fairness in resource puncturing

The data rate of each eMBB UE must be examined as a condition to ensure fairness between eMBB UEs. A Pareto distribution is used to model uRLLC traffic. The probability limitation for issue (2.20) is obtained from the Cumulative Distribution Function (CDF) of the random variable λ that can be converted into a reliable version (Pandey et al., 2019). The optimization process can be simplified using a reasonably simple CDF result. Below is a representation of the Pareto distribution CDF:

$$F_\lambda(x) = \begin{cases} 1 - \left(\frac{x_e}{x}\right)^\epsilon & x \geq x_e \\ 0 & x < x_e \end{cases} \tag{2.20}$$

Where x_e represents the smallest positive value of x and indicates the Pareto distribution's weight, Pareto distributions can be described by a positive metric known as ϵ. The uRLLC traffic reliability limitation can be derived as follows:

$$\begin{aligned} P(E) &= P[\sum_{e \in \mathcal{E}} \eta_e(t) \log_2 (1 + SNR_u) < \lambda(t)] \leq \tau \\ &= 1 - F_\lambda(\sum_{e \in \mathcal{E}} \eta_e(t) \log_2 (1 + SNR_u)) \leq \tau \end{aligned} \tag{2.21}$$

Where $F_\lambda(x)$ is the CDF of λ and $\eta_e(t) \leq \psi_e(t), \forall e \in E$.

Eq. 2.22 illustrates how uRLLC random payload size can be transformed into a predictable metric τ (Alsenwi & Hong, 2018). Thus, the optimization formula is:

$$\begin{aligned} &= \underset{\eta}{\text{maximize}} \sum_{e \in \mathcal{E}} C_e(t) \\ &= \text{Subject to} \sum_{e \in \mathcal{E}} \eta_e(t) \log_2 (1 + SNR_u) \geq F_\lambda^{-1}(1 - \tau) \\ &= \sum_{e \in \mathcal{E}} \eta_e(t) \log_2 (1 + SNR_u) \ \geq F_\lambda^{-1}(1 - \tau) \\ &= \left(\sum_{e \in \mathcal{E}} \eta_e(t) \log_2 (1 + SNR_u) \right)^\epsilon \geq \left(\frac{x_e}{\tau} \right)^\epsilon \end{aligned} \tag{2.22}$$

Here, $F_\lambda^{-1}(1 - \tau)$ is the reverse CDF of uRLLC load to check the predictability factor and assure delivery of the uRLLC services:

$$(\psi_e(t) - \eta_e(t)) \log_2 (1 + SNR_e) \geq \rho \ \forall e \in E \tag{2.23}$$

Each eMBB UE's minimum data rate is ρ, and E illustrates the allocated resource blocks at time slot t.

2.4.3 EDQAS packet control mechanism

The last part of the EDQAS algorithm which is called LDI is modeled to optimize the uRLLC traffic requirements while maintaining fairness across cumulative eMBB UE throughput. In light of the analysis in the previous subsection on the EDC mechanism, it is evident that enforcing efficient priority rules on different service types helps minimize latency on different services. As traffic is variable, it is challenging to achieve low latency at a high data rate level. In light of the analysis in the previous subsection on the EDC mechanism, it is evident that enforcing efficient priority rules on different service types helps minimize latency on different services. Further, scheduling flows that enhance eMBB service throughput is beneficial as long as the overall delay stays the same. It is proposed that the LDI mechanism grant flows rejected by EDC the option of being rescheduled. The SNR of the designated uRLLC service is used to determine the channel condition. Better channel condition means more data is transferred across this RB, leading to improved performance. The payload size of this uRLLC traffic that must be communicated during the current time slot is the problem to be resolved.

The profit-by-weight ratio must be examined to select a resource block with the most significant profit. Puncturing and using the portion of the RB with the most increased profit-by-weight ratio is used as a solution. The TTI length for uRLLC is calculated as a percentage of the RB size, defined as 2, 4 or 7 OFDM symbols.

$$(D_{\text{size}} = D_{\text{size}} - r_u(k)) \tag{2.24}$$

The time execution in recommended EDQAS mechanism is O(N log N) and is appropriate for deployment.

Three threshold indexes are used to describe the packet control diagram. This method counts the number of packets awaiting scheduled $(\forall j \in k)$. A slight delay increase should be applied to packet j to ensure a stable relationship between delay and throughput. The minimum increase in delay D_{inc} is determined as:

$$D_{inc} = D_{HOL,k} - D_{HOL,j} \tag{2.25}$$

Traffic j increases overall delay by a small amount if D_{inc} is less than $I[min(D_{inc})]$. The $I[min(D_{inc})]$ is adjusted by D_{inc} and the following condition should be checked.

The EDQAS scheduler utilizes the symbol μ to express the current throughput for user j relies on the channel condition. Equation (2.25) represents this problem by employing Shannon's capacity to define achievable data rates:

$$\mu = BW_i \log_2 (1 + SNR_i) \tag{2.26}$$

UE average throughput with j or $\overline{\mu}$ is calculated as follows:

$$\overline{\mu} = (1 - \kappa).\overline{\mu} + \mu.\kappa \tag{2.27}$$

Here κ value is set to 0.35 refers to the steady factor, and μ reflects the actual data rate on packet j. Now, the throughput reduction Thr_{dec} induced by packet j is computed as:

$$Thr_{dec} = \overline{\mu}(l) - \overline{\mu}(j) \tag{2.28}$$

If Thr_{dec} is lower than $I[min(Thr_{dec})]$ as the minimum throughput reduction index, then j reduces the total throughput by a small amount. The value of $I[min(Thr_{dec})]$ is then updated by Thr_{dec} here, and packet j is designated to be verified by the next constraint.

Herein the drop increase index packet is defined based on (Madi et al., 2018):

$$Dr_{inc} = 128.1 + 37.6logR \tag{2.29}$$

Here R is the eNB-UE distance in meters. If Dr_{inc} is smaller than $I[min(Dr_{inc})]$, then j increases the total drop ratio with a minor portion and $I[min(Dr_{inc})]$ revised to Dr_{inc}. In the final step, when the packet j meets the above conditions, it is determined for transmission via RB allocation.

Note that the three updated variables $I[min(D_{inc})]$, $I[min(Thr_{dec})]$, and $I[min(Dr_{inc})]$ prevent service j of being scheduled unless it achieves the least degradation in delay, throughput and packet loss ratio $\forall j \in k$. If packet j cannot reach the minimal D_{inc}, it would be measured by the remaining flows of the list k. Eventually, if packet j cannot fulfill the constraints mentioned above, it will be eliminated from the MAC layer. Algorithm 2.2 summarizes the proposed algorithm for EDQAS. In the next chapter, we discuss the implementation of EDQAS in a simulator and afterward study its performance.

Algorithm 2.2 EDQAS Algorithm

```
Initialization;
for TTI = 1 to N_TTI do
    Select Packet && Analyze packet size to define packet type && Compute QoS
      analysis to assign priority weight based on EDC algorithm;
    if packet has best M[j][y] based on EDC criteria then
        if Incoming uRLLC traffic then
            PRB allocation to packet;
            Packet = packet + 1;
        else
            Counting remained packets to schedule eMBB, mMTC;
            while (any (D_size > 0) && any (P < Number of Mini-Slots) ) do
                Choosing a uRLLC service
                Using Eq. 2.1 to calculate eMBB data rates at each RB and save in R
                Determine the SNR_u and R to record the result in R_a
                Sort R_a from highest to lowest
                if (P(i) < num of mini-slots) then
                    Puncture RB at R_a(i)
                    P = P + 2
                    i = i + 1;
                end if
                Update D_size using equation 2.24
                if (i > Number of RBs) then
                    i = 1;
                end if
            end while
            Compute delay increment within packet scheduling;
            if ((Packet cause minimum increase on delay?) &&(Packet cause minimum
              decrease on throughput?) &&(Packet cause minimum increase on drop
              ratio?)) then
                Update QoS criterias;
                PRB allocation to packet;
                Packet = packet + 1;
            else
                Discard the packet;
                Packet = packet + 1;
            end if
        end if
    else
        Discard the packet; Packet = packet + 1;
    end if
end for
```

CHAPTER 3

5G AIR SIMULATOR

3.1 Introduction

New technologies have consistently been designed, evaluated, and optimized with the help of computer simulation, allowing for cheaper and faster research than physical prototypes. The 5G-air-simulator represents significant progress in this area. An open source tool that simulates 5G scenarios including broadcasting, improved random access, and NB-IoT. Further, It allows us to reproduce the different forms of traffic that 5G supports to analyze how their specific requirements are met.

3.2 Overview of 3GPP New Radio and 5G Services

Between 5G and 4G, there are significant distinctions. Data traffic for 4G is generated mainly by individuals. However, 5G also involves services that require no human interaction, such as interconnected automobiles and IoT appliances. Even data sharing based on human interaction could have different needs than LTE. For instance, in virtual reality, the importance of extremely low latency is greater than the need to optimize throughput.

In 5G NR, a new air interface will coexist with the 4G LTE network. Nonetheless, 5G goes beyond simply being able to support significantly faster data rates and offering consumers more capacity than 4G. High-frequency bands, TDD, and low latency support are some of the characteristics of NR technology. Additionally, 5G NR must deliver various services across various devices with different performance and latency requirements to meet 5G expectations. A simulation platform that supports 3GPP NR to generate a traffic combination of uRLLC, eMBB, and mMTC scenarios is required. The 5G-air-simulator was the best choice for this work. There are multiple simulators like NS3, but 5G air simulator is the best fit for this simulation since it is difficult to generate different KPIs ("Delay", "Goodput", and "PLR") for different traffic types.

However, the 5G air simulator provides the necessary functions to generate the specified results simultaneously.

3.3 Overview of the 5G-air-simulator

The 5G-air-simulator is an expansion of the LTEsim network tool to provide a 5G interface; its source code is freely accessible at (Martiradonna *et al.* (2020)). The simulation framework incorporates the key features of the LTE-Sim tool to model multi-user, single-cell, single-cell with interference, multi-cell, and heterogeneous network. Additionally, since this tool is open-source and modular, numerous technological components have been added to provide opportunities for developing research areas like cellular-connected drones and Network Slicing.

3.4 Simulation implementation

A performance analysis employing numerical simulations is conducted in comparison with existing scheduling techniques designed for different RT traffic to demonstrate the efficiency of EDQAS as a downlink MAC scheduler. The relevant scheduling algorithms are PF, MLWDF, EXP-PF, and it is worth mentioning that in this thesis, the proposed **EDQAS** algorithm is implemented in the 5G air simulator as a new MAC algorithm scheduler for the first time. The evaluation emphasizes the ability of the schedulers to reduce latency as the primary goal of this thesis associated with sustaining an acceptable QoS level. In the following, we take a look at PF, MLWDF, and EXP-PF the alternate MAC scheduling algorithms:

Proportional Fairness (PF)

Proportional fairness is a resource scheduling concentrated on fairness. It is focused on keeping the right balance between two opposing goals: Attempting to optimize overall network throughput while ensuring that all customers receive the service. The PF algorithm is considered ineffective in real-time applications and penalizes delay-sensitive UEs (Mamman et al., 2019).

Modified Largest Weighted Delay First (M-LWDF)

In real-time applications, the M-LWDF approach is recommended to guarantee Quality of Service (QoS). M-LWDF provides services to users based on the channel status and the status of each user's queue. Using this algorithm, a more significant portion of users can achieve acceptable QoS and ensure a minimum throughput. However, it strangles non-real-time applications and fails to meet 3GPP QoS criteria (Lawal et al., 2017).

Exponential Proportional Fairness (EXP-PF)

In addition to providing throughput for non-real-time applications, this algorithm also guarantees the quality of service for real-time applications. EXP rule and PF are the two strategies the EXP/PF use. While the PF distributes resources to non-real-time applications, the EXP allocates resources to real-time apps. It ensures a good system throughput and offers QoS to real-time applications. Nevertheless, it is improper to non-real-time applications when there are many users (Lawal *et al.* (2017); Mamman *et al.* (2019)).

3.4.1 Simulation tool

3.4.1.1 Core

An overview of 5G air simulator features is provided below. Figure 3.1 shows the critical components of the created tool. Simulators are built on their core, which contains all processes necessary for implementing and handling protocol stacks. 5G-air-simulator organized as an event-driven application: The Calendar class stores events, methods, objects, and parameters. Two more significant classes are simulators and frame managers. Events are added to the calendar, and simulations are started and stopped. The Frame-manager monitors the passage of time and increments the counters associated with frames and sub-frames. Events associated with frames and subframes are scheduled using a fixed frame structure.

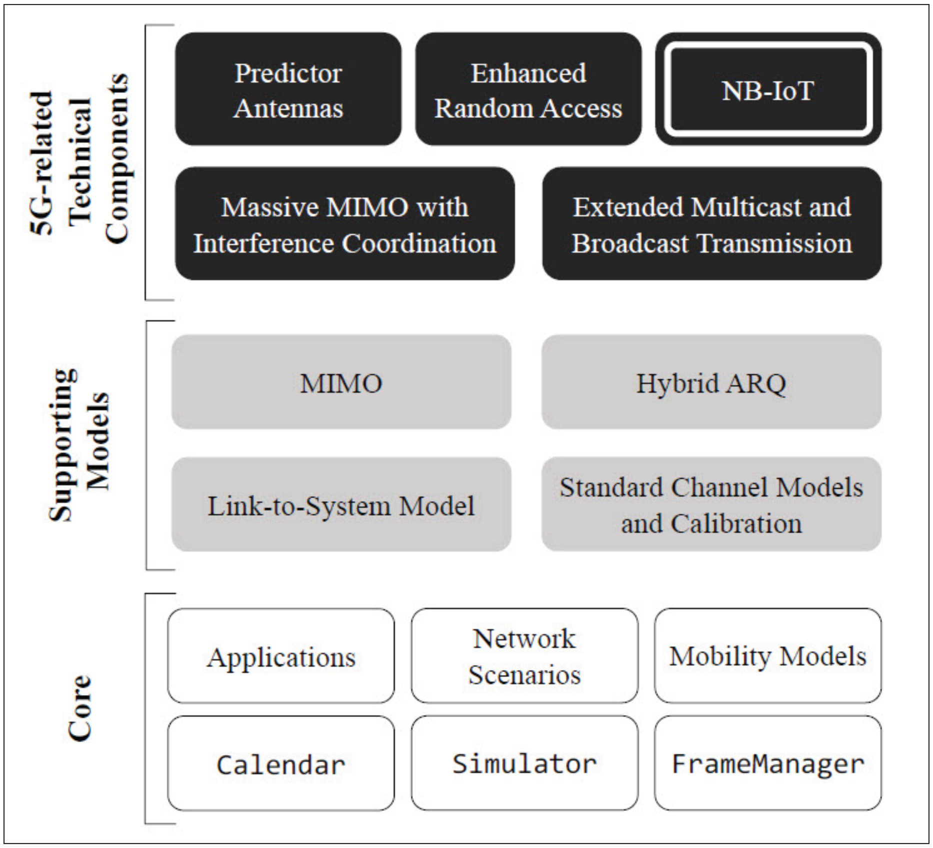

Figure 3.1 5G air simulator core
Taken from Martiradonna *et al.* (2020)

3.4.1.2 Protocol Stack

An application model produces data packets that are transmitted through the protocol stack transceiver and processed by the receiver. The "Trace-Based" approach aims to capture the traffic produced by streaming videos. The model includes several traces made from an actual video frame and provides the flow rate of each frame. A new trace is initiated upon reaching the end of the previous trace. I chose this traffic model since the largest packet sizes should be allocated for eMBB corresponds to a 128kb Trace-Based model.

A "VoIP" model for voice traffic uses a G.729 codec model to generate packets at a constant rate and size when the connection is in active mode. During an idle mode, no packets are generated. Throughout human speech, the active state can switch to the passive form anytime. I chose this traffic type for uRLLC services since it creates no data in idle mode. Afterward, the network experiences no further delay. Additionally, the needed packet size for uRLLC services like tel-medicine or tel-surgery corresponds to a 32kb VOIP model.

In general, the constant bit rate is considered to be the most basic model. At predefined periods, it creates packets of a fixed size. The model can describe applications that send data regularly, such as remote sensors that receive reports from distant locations. I chose this type of traffic for mMTC applications like smart homes and drones since they will generate sporadic traffic between many geographically spread devices to provide long-distance connectivity through periodic reports from remote sensors.

Moreover, the 5G-air-simulator provides additional user- and control-plane protocol stack capabilities. Each device implements a Protocol Stack class instance for this purpose, which in turn comprises entities for MAC, RLC, PDCP, and Application. Packets originating from the higher layer are primarily handled by the PDCP Entity's header compression and enqueued into the MAC Entity (Martiradonna *et al.* (2020)).

3.4.1.3 Network Deployments

In Figure 3.2, a variety of scenarios are illustrated in the 5G-air-simulator. As its name implies, the simplest is SingleCell consists of just a single cell and a single base station, serving a limited number of users without any interference. By utilizing this design, remote sites or ideal conditions can be evaluated for peak throughput or coverage.

However, In the "**SingleCellWithInterference**" scenario, a more accurate configuration is developed to meet the EDQAS algorithm requirement for the test scenario because this is the only comprehensive scenario to generate delay, packet drop ratio, and goodput results simultaneously. Despite the presence of a single omnidirectional base station, other unusable ones surround it. All other gNodeBs still cause inter-cell interference in all radio channels, modeled as an always-on transmission. Therefore, neighboring cells surrounding the primary cell affect performance metrics.

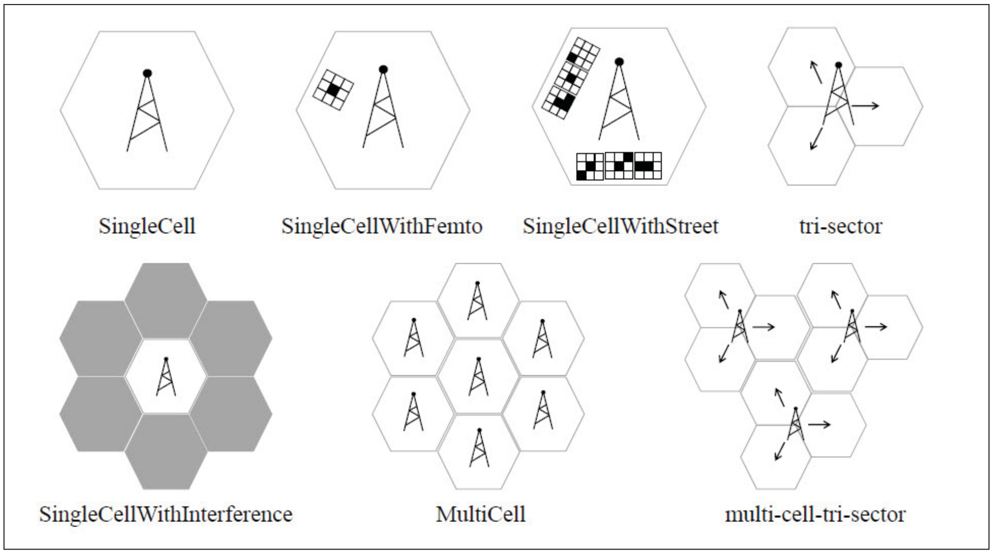

Figure 3.2 Available Network deployments
Taken from Martiradonna *et al.* (2020)

A macro-cell gNB is installed in the middle of the 5G cell area. The gNB establishes a direct link with the users.

3.4.1.4 Link Adaptation

A link adaptation determines which MCS best fits the channel quality of each user. Alternatively, an excessive modulation level will transfer more data and significantly increase physical layer errors, so the benefit is lost. On the other hand, a too-low modulation level would cause a decrease in speed without any appreciable benefit. Users calculate SINR values, which are derived by evaluating the channel quality with Channel Quality Indicators (CQIs). The gNodeB is responsible for translating the CQIs to MCS indexes. Transport Block Size defines as the amount of payload transferred to the user by the MAC layer using MCS indexes.

3.4.1.5 Calibrated Link-to-System Model

The link-to-system model calculates radio transmission efficiency while accounting for propagation and interference. Modeling an extensive system with a link-to-system approach offers a condensed description that is still accurate. Figure 3.3 shows the model for the 5G air simulator. Signal attenuation due to path loss propagation and fast fading is estimated using the Urban Macro-cell scenario. Fast fading simulates Transmission Time Interval (TTI) characteristics, including delays, powers, arrivals, and departures.

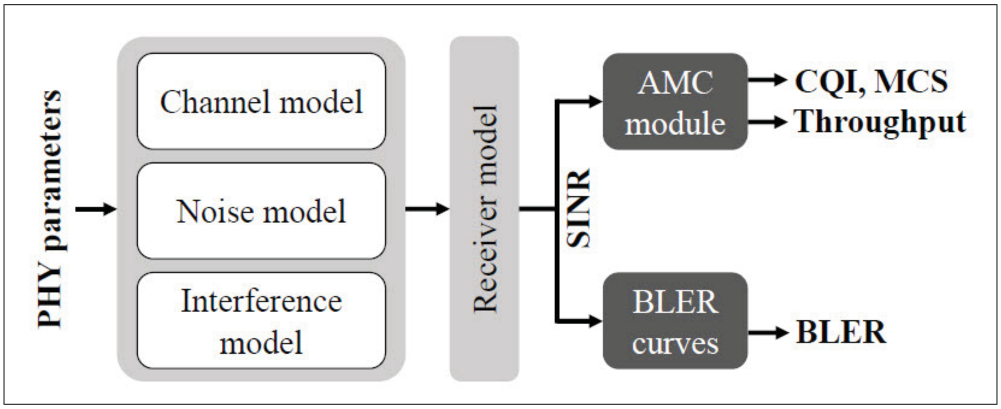

Figure 3.3 link-to-system model
Taken from Martiradonna *et al.* (2020)

3.4.1.6 Simulation Tracing

Figure. 3.4, illustrates the content of a 5G air simulator trace. Each line gives information about the event; for example, packets transmitted, received, or discarded are indicated by rows beginning with TX, RX, and DROP, respectively. The PHY RX line describes a physical layer receiving event, and RANDOM ACCESS describes random access. An application type is defined in the second field of packet lines. "Id" uniquely determines the packet, and "B" maps it to the bearer. "SRC" (source) and "DST" (destination) define the sending and receiving nodes. The received packet delay is indicated by "D". An event's time is reported by "T".

Performance metrics are extracted from the output of each simulation run. Here are a few examples of key performance indicators.

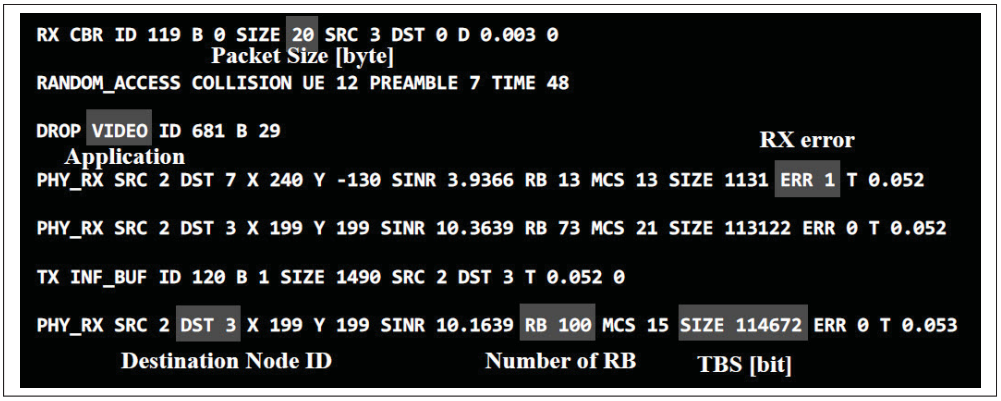

Figure 3.4 Text trace simulation sample
Taken from Martiradonna *et al.* (2020)

3.4.1.7 Delay

Using all the lines beginning with RX, the 14th position is expressed in seconds. In this case, averaging all the values is sufficient.

3.4.1.8 Goodput

At the eighth position, add up the values beginning with RX, indicating the cumulative size transmitted of the application data. To determine the goodput in bps, multiply the sum by eight and divide by the simulation duration.

3.4.1.9 Packet Loss Ratio (PLR)

At the application layer, PLR is computed as the proportion of dropped packets to all sent packets. Transmitted packets are TX, and received packets are RX.

3.4.2 Test Scenario

As seen in figure .3.5, each user receives "6" uRLLC, "2" eMBB, and "2" mMTC from the gNB (service provider) in downlink simultaneously. The 5G air interface was simulated and examined using the 5G-air-simulator. Additionally, the tool described in this study is modular and open-source; new technological components may be included to open up future research simulations which has allowed us to add the EDQAS scheduler.

Scenario	Simulation Time [s]	Real Time [s]	Real Time/ Simulation Time
SingleCell	25	30	1.20
SingleCellWithInterference	46	27	0.59
MultiCell	100	711	7.11
SingleCellWithFemto	30	547	18.23
f5g-uc1 (mMIMO) - 104 UE (Rural)	10	3714	371.40
f5g-uc1 (mMIMO) - 150 UE (Suburban)	10	4510	451.00
f5g-uc1 (mMIMO) - 104 UE (Urban)	10	3808	380.80
f5g-uc6 (Extended Multicast)	10	2028	202.80
f5g-uc2 (High-speed)	6	1397	232.83
MMC1 (Enhanced Random Access) - 1560 UE	11	1	0.09
MMC1 (Enhanced Random Access) - 156000 UE	15	3357	223.80
nbCell (NB-IoT) - 1200 UE	150	55	0.37
nbCell (NB-IoT) - 3600 UE	150	993	6.62

Figure 3.5 5G-air-simulator configuration
Taken from Martiradonna *et al.* (2020)

3.4.3 Traffic Models

The simulation scenario comprises three distinct traffic types: uRLLC, eMBB, and mMTC. 60% of users utilize uRLLC, 20% incorporate eMBB, and the remaining 20% use the mMTC application.

A trace-based producer implements the eMBB application. It transmits packets from actual trace files (Seeling et al., 2011). The H.264 standard encodes eMBB data sequences for creating variable bit rate streams of 128 Kb. The uRLLC and mMTC applications run lightweight traffic based on G.729, an ITU standard (Salami et al., 1998). Modeling the uRLLC is typically based on an ON/OFF Markov chain with a mean interval of 4 seconds for the ON period. In comparison, OFF period has an upper limit of 6.9 seconds and an average value of 3 seconds (Chuah et al., 2002). When the uRLLC source is ON, it transmits 20 bytes every 20 ms, and the mMTC application source transmits packets of 28 bytes every 20 ms. Each uRLLC and mMTC packet, has a 12 bytes header. In contrast, no transmission takes place during the OFF period.

Table 3.1 Description of simulation parameters

Parameters	**Description**
Bandwidth	10 MHz
CELS=1	NUMBER OF CELLS
FILE="Sim"	OUTPUT FILE NAME
FILENAME="Multi"	SIMULATION TYPE NAME
Frame structure	TDD
gNBs in cell	1 gNB
INTERVAL=10	users growth increment
MINUSERS=10	Minimum of users
MAXUSERS=160	Maximum of users
Max delay bound	100 ms
NUMSIM=1	Number of simulations
Number of Users	10-160 with a period of 10 Users
Number of uRLLC services per user	6
Number of eMBB services per user	2
Number of mMTC services per user	2
PRBs allocation time	1 ms
Radius in Km	1Km)
User speed	3km/h
Users applications rates	128 kbps eMBB, 32 kbps uRLLC, 40 kbps mMTC

Specifically, the performance evaluation research is to the ability of EDQAS to minimize latency when transferring different traffic types. Furthermore, stability in keeping an acceptable QoS level under large network loads is a possible area to be investigated. These assessment criteria can help us understand how to apply the proposed strategy to practical mobile applications. Table 3.1 includes additional descriptions of simulation parameters. The carrier frequency range of 2.1 GHz is used at the physical layer (2110-2170 MHz). Each subcarrier has a spacing of 30 kHz and has a power transmission of 43 dBm. This work integrates four separate modules to meet 5G requirements. Note that the maximum number of users corresponds to the saturation of the capacity of the cell. In the next section, we look at the results of our simulation.

CHAPTER 4

NUMERICAL RESULTS AND ANALYSIS

4.1 Performance Evaluation Results

Figures 4.1, 4.2, and 4.3 illustrate the end-to-end delay values. An application layer sends the traffic from the traffic source across the channel to the user. The delay is described in this way. D_{E2E} can be calculated as follows:

$$D_{E2E} = D_{queue} + D_{prop} + D_{trans} \tag{4.1}$$

Note that D_{queue} represents the queue delay at the MAC/RLC layer, and it typically has a significant influence on D_{E2E} traffic, particularly bursty traffic. Furthermore, D_{prop} and D_{trans} are propagation and transmission delays produced by the wireless medium between gNB and the user. Figure 4.1 presents the average D_{E2E} for uRLLC traffic. Compared to PF, MLWDF, and EXP-PF, EDQAS obtained the best minimal latency across rising traffic load, which indicates that EDQAS is the most effective algorithm in a high overload situation.

This optimization is enforced by the efficient delay-based decision established in the EDC technique. In brief, by using the EDQAS algorithm uRLLC is prioritized if the buffer size and channel quality condition are sufficiently large. EDQAS demonstrates a slight increase in delay compared to EXP-PF for 90 UEs, then a reduced D_{E2E} pattern is observed by EDQAS when the cell reaches maximum utilization in 160 UEs. Whereas EXP-PF depicts an appropriate D_{E2E} behavior compared to MLWDF, due to a delay threshold (D_{max}) that prohibits excessive data transmission during congestion periods.

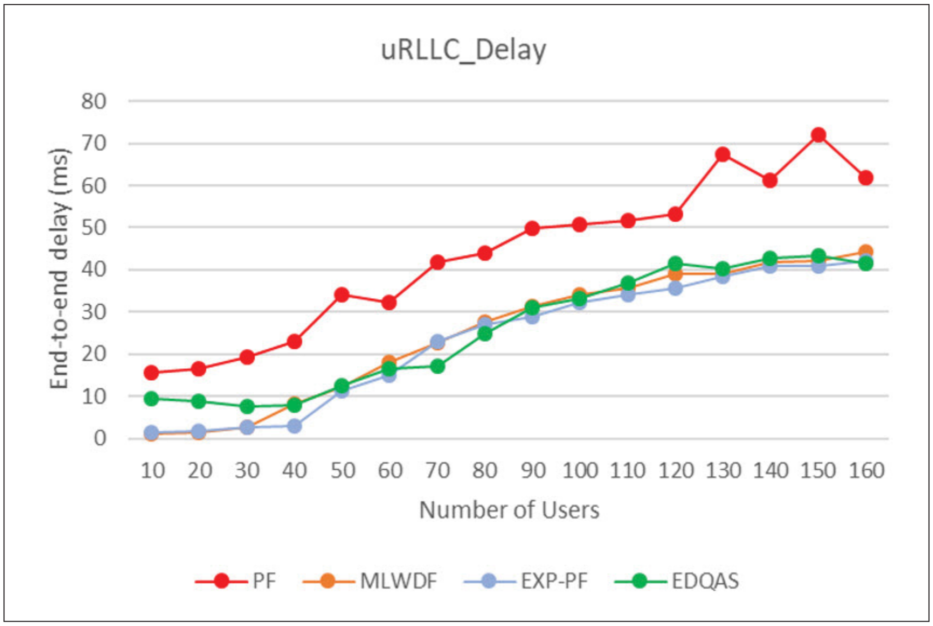

Figure 4.1 uRLLC Delay

The result in figure 4.2 for eMBB traffic supports the previous comments about minimizing delay. The analysis revealed that EDQAS reduces delay considerably against EXP-PF in the case of 80 UEs.

The Ex-PF and MLWDF algorithms are designed to comply with the QoS criteria for real-time applications based on delay thresholds and packet loss ratios. The gNB selects users with higher CQI for higher MCS because poor channel quality results in lower QoS. Nevertheless, EDQAS performance is better than other schedulers because of its use of resource puncturing, contributing to lower latency.

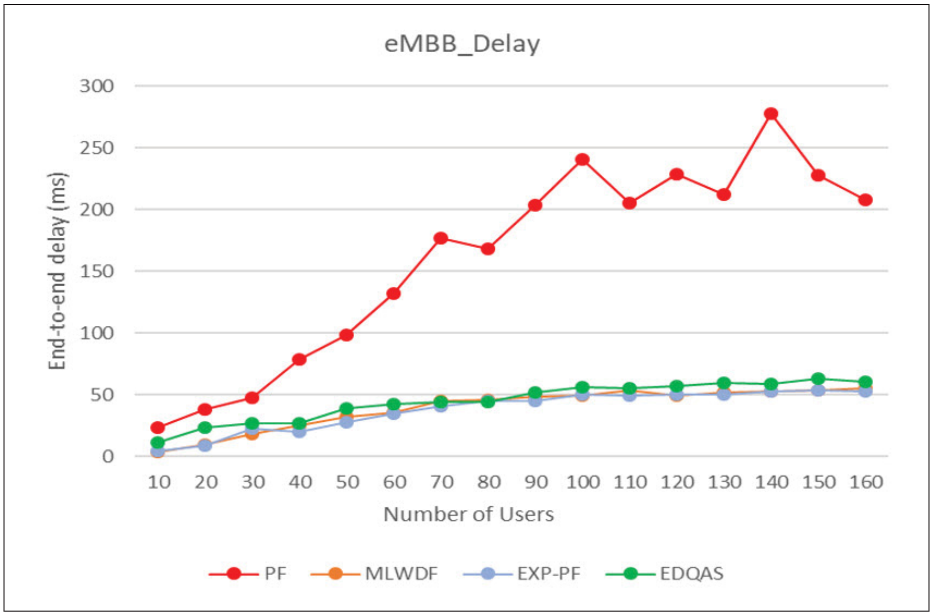

Figure 4.2 eMBB Delay

Even if EDQAS can achieve acceptable latency across the presented load levels for mMTC applications, sustaining a high throughput and low packet loss ratio is challenging. As illustrated in figure 4.3, EDQAS outperforms all other algorithms and demonstrates that it is a relatively stable latency pattern for mMTC traffic.

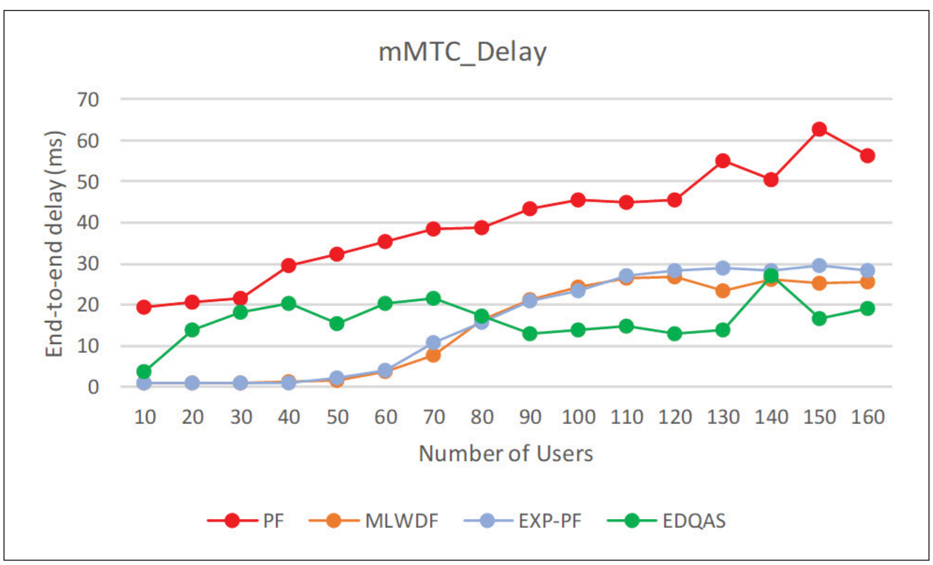

Figure 4.3 mMTC Delay

A discussion of the performance of EDQAS in achieving high goodput for uRLLC traffic is also presented. According to (Madi *et al.*, 2018), goodput is the number of usable bits successfully transmitted from a traffic source. Typically, uRLLC traffic poses little stress to the MAC scheduler since it is ultralight. The goodput results for uRLCC traffic in figure 4.4 demonstrate the priority metric rule for uRLLC packet in EDQAS to increase goodput. The scheduling decision is primarily channel-aware for services transmitted to high CQI UEs.

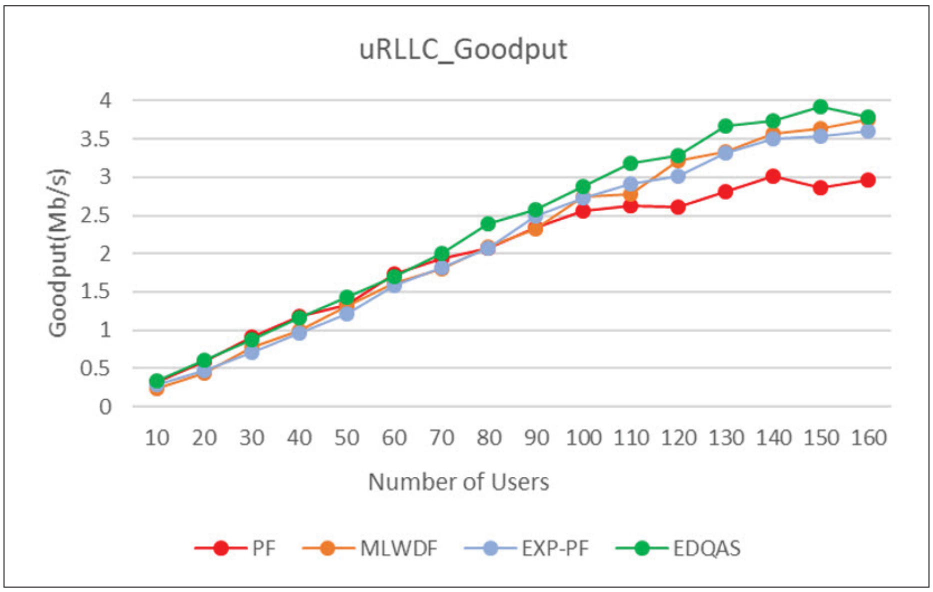

Figure 4.4 uRLLC Goodput

The results in figure 4.5 illustrate an excellent trend of EDQAS on eMBB goodput. The eMBB and uRLLC traffic are dynamically allocated to optimize eMBB goodput while meeting uRLLC user requirements. The primary contribution of this study was that the quality and reliability level of uRLLC substantially influences eMBB users. In particular, eMBB QoS is based on ensuring a high data rate for the UE. EDQAS improves goodput noticeably by adopting the slightest delay increase technique in which packets are scheduled based on respective data rates.

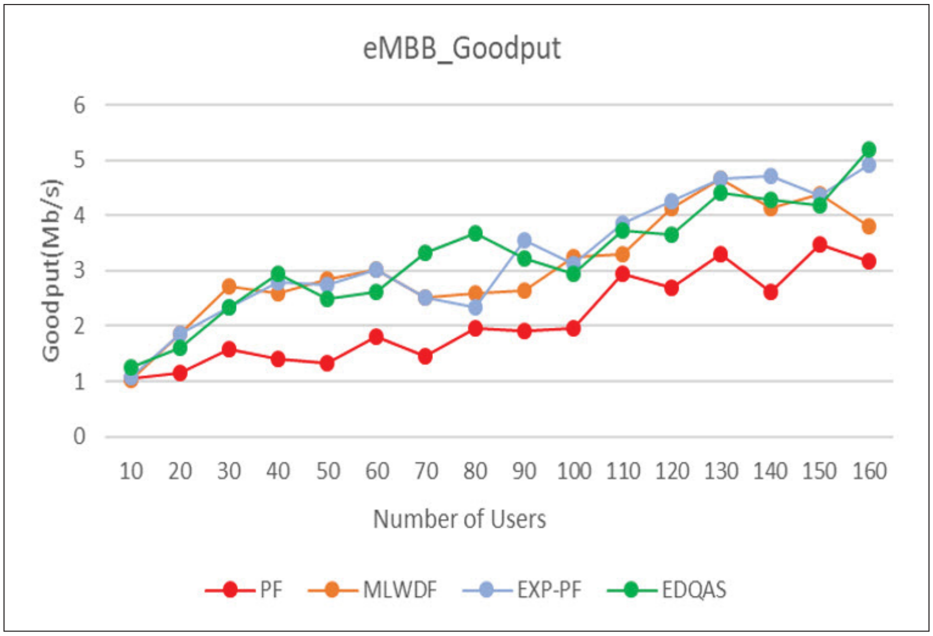

Figure 4.5 eMBB Goodput

Obviously EDQAS cannot obtain good results for mMTC goodput, unlike uRLLC latency and enhancing eMBB data rate. Hence, acquired results of goodput and packet loss ratio for mMTC application might be negligible, as illustrated in figure 4.6 and 4.9. For more justification, this research is focused on improving QoS level, including lower latency, lower packet loss ratio, and better goodput for **uRLLC** and **emBB** services, not mMTC. The results of mMTC traffic are only mentioned to make this study a comprehensive version of examining different traffic types of 5G features with the 5G air simulator.

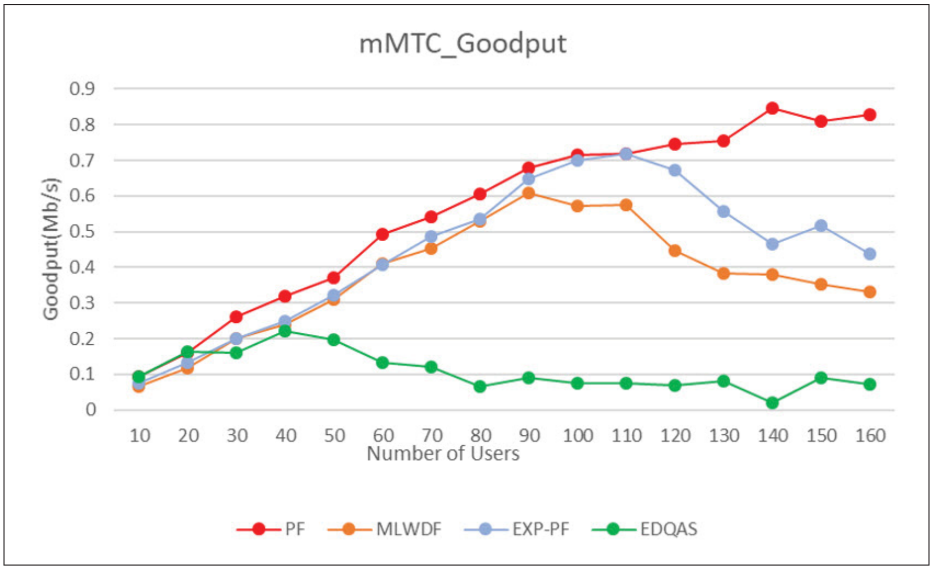

Figure 4.6 mMTC Goodput

We now study the packet-dropping ratio for various types of traffic. This study estimates data dropping as a percentage of the total amount of data transferred at the RLC and MAC levels. Maintaining low data drop ratio demonstrates that EDQAS is able to respond under high traffic network. For uRLLC traffic, providing lowest packet loss ratio has a high importance. Figure 4.7 illustrates the effects of data drop ratio on uRLLC traffic and EDQAS offers the best result to achieve minimal packet loss, as shown in the statistics. The EDQAS method can ensure the same constrained latency even in a momentary channel disruption scenario when a flow cannot transfer the proportion of data computed by EDQAS until the end of the current frame. Each TTI is impacted by the amount of data in the transmission queue. In figure 4.7, when a significant decrease in channel quality occurs around 40-60 users and prevents the transmission of the scheduled date, EDQAS will process the waiting packets, scheduling them in the following frame and allowing a larger amount of data to be delivered.

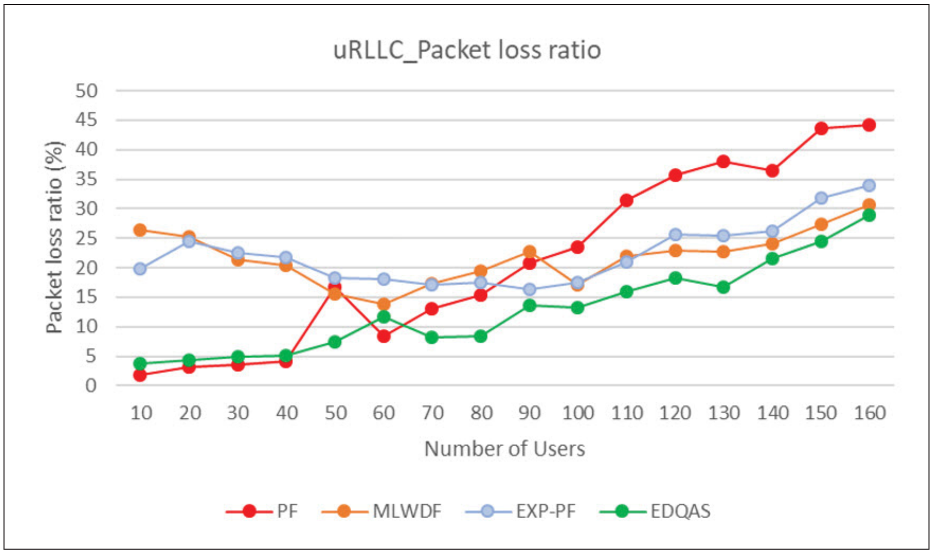

Figure 4.7 uRLLC Packet loss ratio

In figure 4.8, the efficiency of QoS-aware methods is shown, with EDQAS outperforming other algorithms in terms of packet loss ratio for eMBB traffic. This study highlights that EDQAS achieves the lowest packet loss ratio and smallest delays, but at the cost of less resources for mMTC traffic, as we can see in figure 4.6.

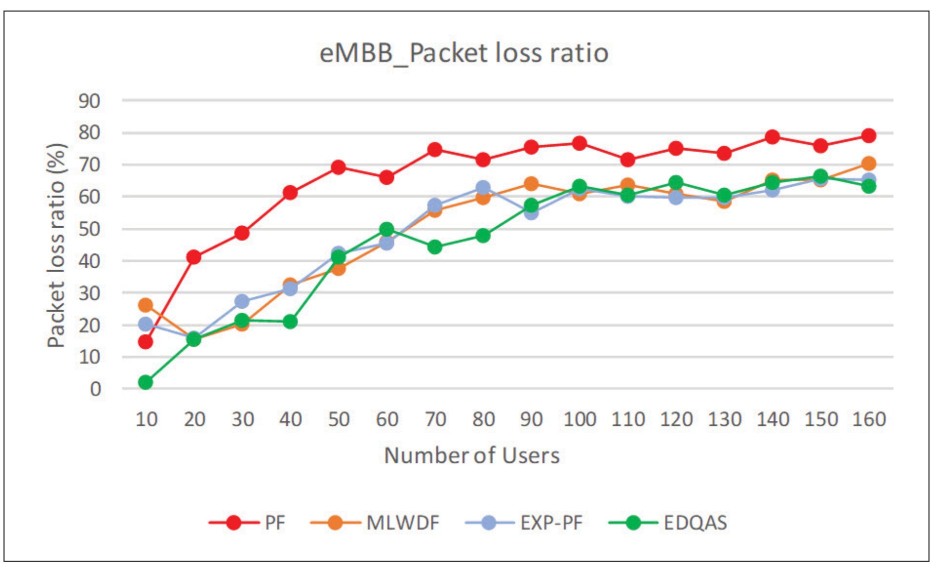

Figure 4.8 eMBB Packet loss ratio

There are some bottlenecks in 5G services. A large part of the difficulties associated with mMTC support can be attributed to congestion in the cellular random access channel due to multiple Machine-Type Communications sending access requests simultaneously in the absence of sufficient preamble resources (Zhan et al., 2021).

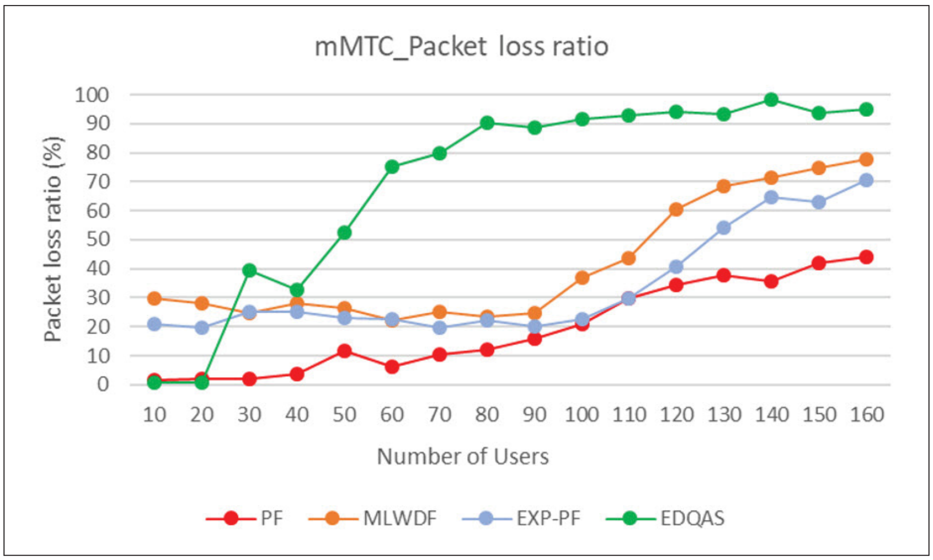

Figure 4.9 mMTC Packet loss ratio

Figures 4.7, 4.8, and 4.9 demonstrated how EDQAS would evaluate waiting packets, scheduling them in the next frame and specifying a greater quantity of data to be sent when a severe reduction in channel quality hinders the transmission of the planned data for a period of 40 to 60 number of users.

4.2 Link Utilization

Link utilization is the amount of traffic that is transferred through a link represented as a proportion of the overall link capacity shown in figure 4.10. In this study, the resource scheduling algorithm employs non-orthogonal slicing to boost spectrum efficiency, but it also interferes with other types of traffic, particularly mMTC applications. It is worth mentioning that the link reaches saturation at 160 UEs.

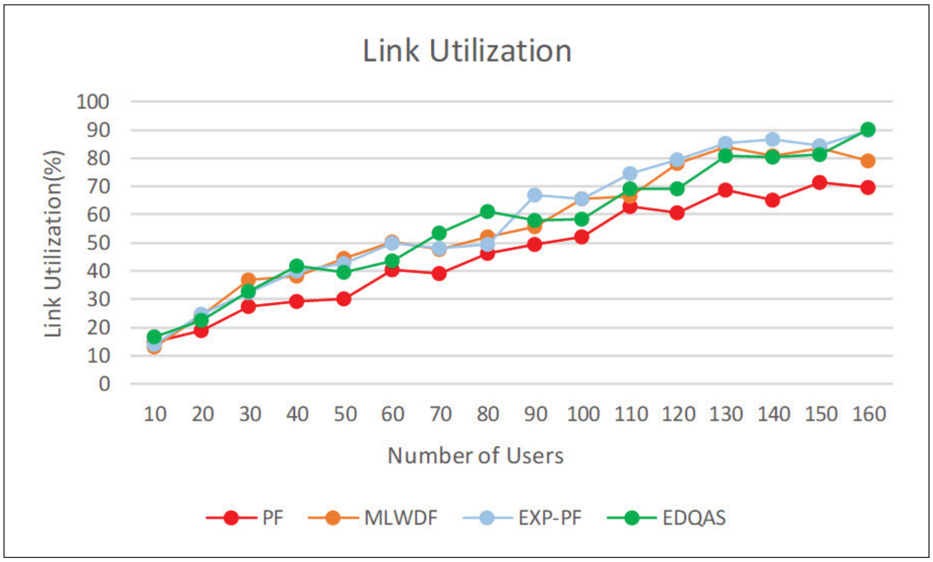

Figure 4.10 Link Utilization

4.3 Overall Delay

The overall EDQAS delay for all 5G traffic types is represented in figure 4.11. The analysis indicated that the EDQAS algorithm performs better for uRLCC applications with 70 users or less, while it plays better for mMTC applications with higher traffic loads.

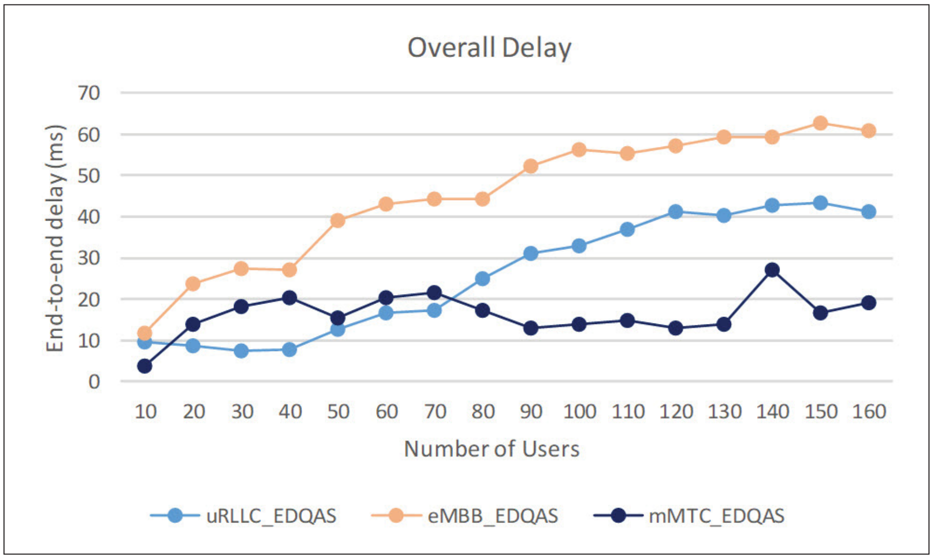

Figure 4.11 Overall Delay

4.4 Overall Goodput

Results in figure 4.12 display a remarkable EDQAS trend on eMBB goodput. Providing high data rates for subscribers is crucial to eMBB users. A slight delay increase method that plans flows according to their data rate would enable EDQAS to increase the goodput rate.

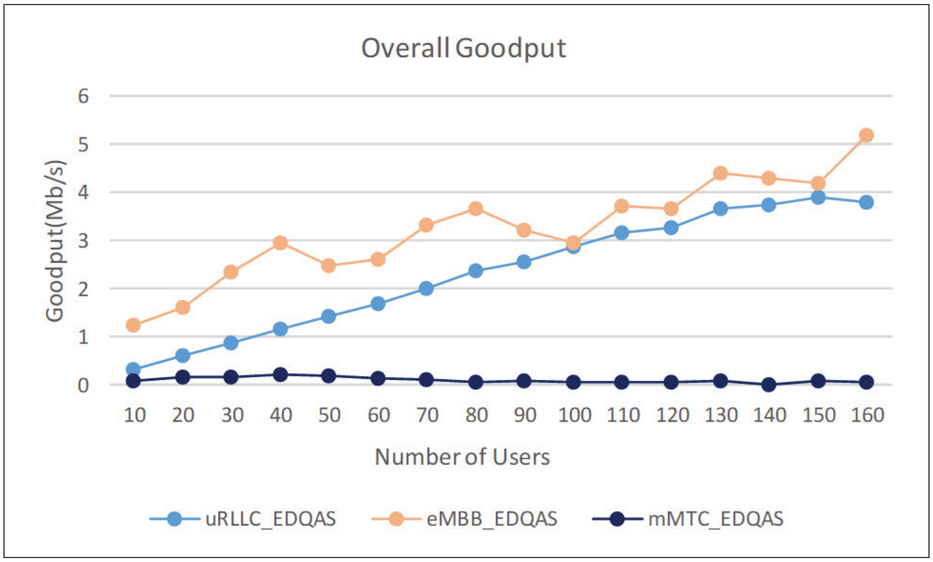

Figure 4.12 Overall Goodput

4.5 Overall Packet Loss Ratio

EDQAS preserved stable and low delay behavior, especially on uRLLC traffic, independent of the increased network load as shown in figure 4.13.

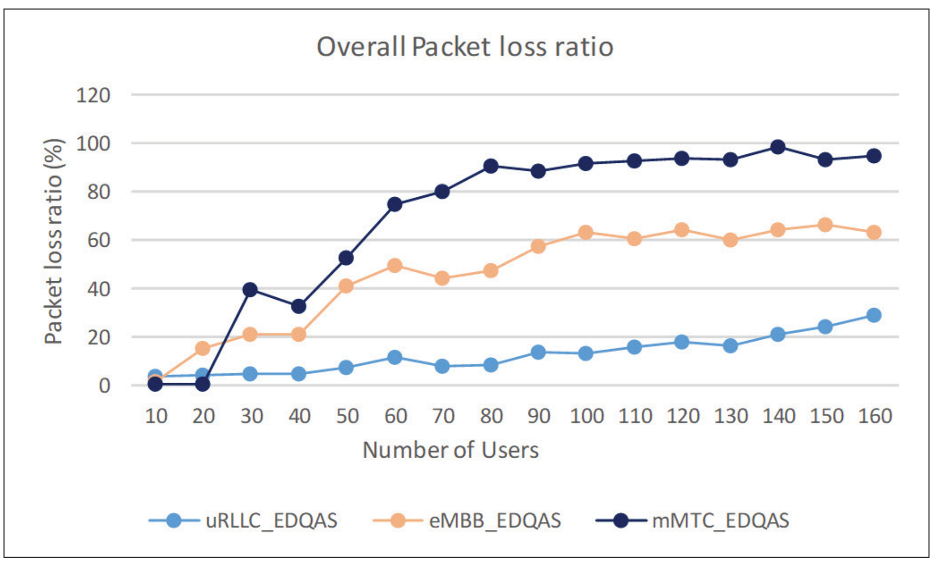

Figure 4.13 Overall Packet loss ratio

Note that the results for the mMTC application may be insignificant. This research aims to improve uRLLC latency and increase the eMBB data rate by utilizing the EDQAS algorithm, resulting in an enhanced level of QoS. This research cannot achieve optimization for all traffic types including uRLLC, eMBB, and mMTC. Some improvements are achieved for uRLLC, and eMBB. On the other hand, mMTC is suffering. However, this study does not examine latency reduction, packet loss ratio reduction, or higher goodput improvements for mMTC applications.

CHAPTER 5

CONCLUSION AND RECOMMENDATIONS

We have proposed EDQAS for optimized resource allocation scheduling in the 5G MAC layer with the aim of reducing uRLLC latency and increasing eMBB data rates.

First the EDQAS algorithm executes the developed efficient delay control (EDC) mechanism to prioritize resource allocation for different services to fulfill delay conditions, especially for uRLLC.

Based on the channel conditions of users, an optimal selection of RBs is proposed in the puncturing phase. The SNR of the selected uRLLC service is used to measure the channel condition. This parameter reflects the impact of puncturing this RB on eMBB UEs' goodput. The resource puncturing part defines the payload size for a uRLLC service in the given time slot. When the uRLLC traffic arrives randomly, the base station punctures previously allocated eMBB resources to serve uRLLC traffic immediately.

Furthermore, the least delay increase mechanism, or LDI, is proposed to schedule services with minimal increase of delay, minimum decrease of goodput and a minimum increase of drop ratio compared to predefined indexes for these KPIs.

Based on simulation results, the proposed EDQAS maintained low delay for high network loads compared to the benchmark scheduling mechanisms. For eMBB a large goodput is a critical factor. The EDQAS increased the goodput rate using the LDI mechanism, which schedules flows based on data rates. Therefore, for a large number of users, EDQAS greatly improved eMBB goodput. EDQAS is sustaining a low and robust delay behavior while giving maximum goodput with minimum data loss on uRLLC and eMBB services, irrespective of the network load. As an efficient and lightweight scheduler, EDQAS is well-suited to mobile networks.

Future Research Work

There are multiple outstanding research areas in resource management for the 5G MAC layer to achieve further performance improvements. The following is a summary of main issues that could be addressed in future research:

1) Further investigation is needed to optimize resource allocation and minimize the degradation of other services, such as mMTC and eMBB.

2) Using **Opus** as another low algorithmic delay coding format for uRLLC services and examining the impact on other QoS parameters, specifically goodput.

3) Managing different radio slices could be accomplished by extending **DownlinkPacketScheduler** and **UplinkPacketScheduler**.

4) These promising results encourage further investigation of EDQAS schemes in 5G **multi-cell-tri-sector** scenario to support mMTC services.

5) **mmWave communications** are under investigation by the 3GPP due to their speed and latency. Adaptive beamforming and large antenna arrays enable highly directional transmissions. There are several features affected by this approach, including cell search, broadcast signaling, and random access. Thus, mmWave 5G systems require considerable research. mmWave communications require that the class ChannelRealization be updated to integrate new channel models and BandwidthManager properly, and AMCModule be extended to accommodate extended bandwidths.

6) Signal processing is viewed as another critical bottleneck for achieving ultra-reliability and low latency. Due to the limitations of current signal processing capabilities, low-complexity algorithms are being researched and utilized. Parallel processing and denser processors in a small electronic device might be helpful to enable the implementation of sophisticated signal-processing methods.

APPENDIX I

THE 5G AIR SIMULATOR

1. Tools installation

This simulator is designed in C++ and incorporates event-driven and object-oriented concepts. It is recommended to run this simulator on Ubuntu OS. The step-by-step instruction is described here.

$ sudo apt install build-essential

Check C compiler version:

$ g++ –version

1.1 Getting 5G-air-simulator

For a 5G-air-simulator, use this script to obtain the preferred folder.

$ git clone https://github.com/telematics-lab/5G-air-simulator.git

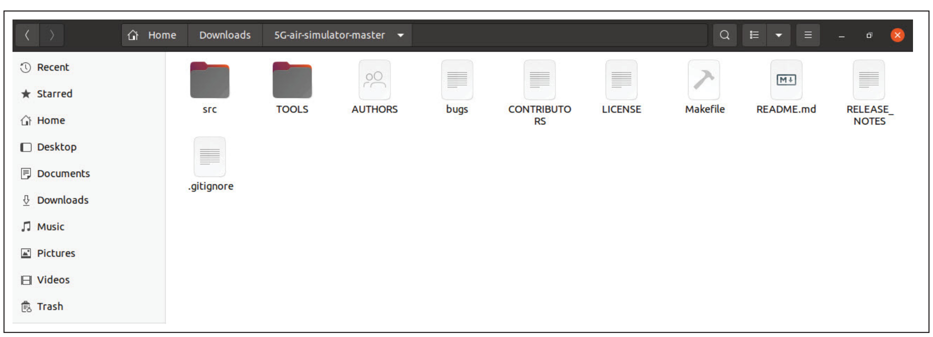

Figure-A I-1 Getting 5G air simulator

1.2 Compiling the 5G air simulator

This command builds the 5G-air-simulator:

$ cd 5G-air-simulator; make

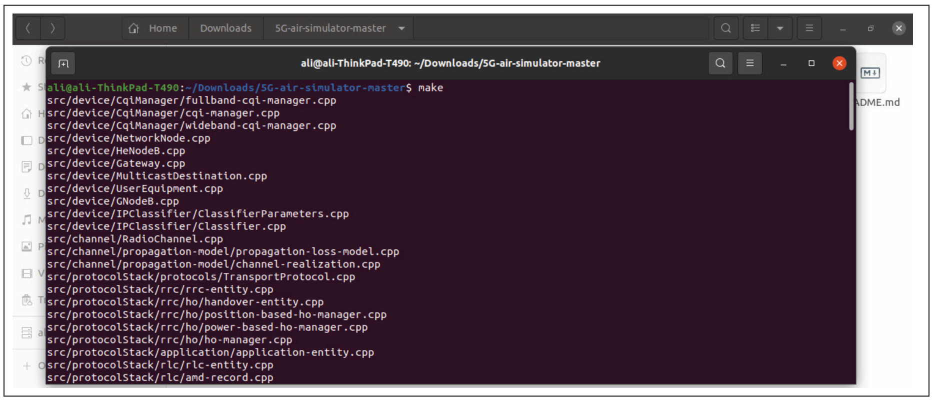

Figure-A I-2 Compiling 5G air simulator

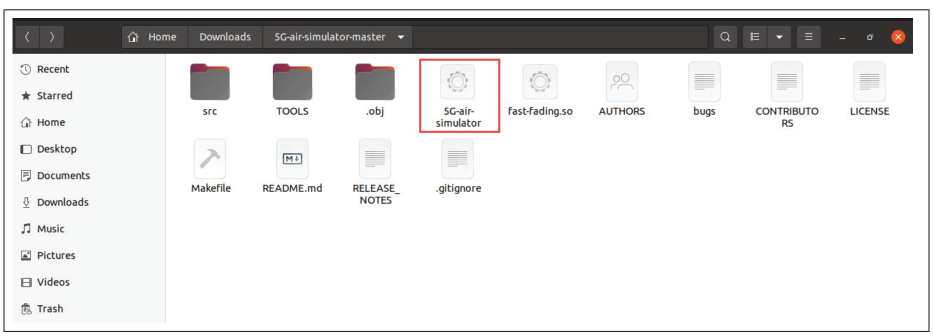

Figure-A I-3 Compiled 5G air simulator

This command clears the project:

$ make clean

1.3 Running the 5G air simulator

To run a simple simulation, use the following command:

$ ./5G-air-simulator Simple

Here are the available scenarios:

$ ./5G-air-simulator -h

1.4 Additional files to the 5G air simulator

It is necessary to create a folder and copy the following shell commands (.sh files) from LTESim into your 5G air simulator folder to determine the delay (delay-comp.sh), goodput (compute-throughput.sh and throughput-comp.sh), packet loss ratio (compute-plr.sh), spectral efficiency (link utilization → specEff.sh) and generate results.

Figure-A I-4 Additional files to the 5G air simulator

The main file is doSim1.sh. This is the file where all the parameters are set up for simulation. This file makes simulations using only these algorithms PF, EXP/PF, and M-LWDF. Then the proposed algorithm should be added. What "doSim1.sh" makes several simulations for each scenario, gets the simulations average, and finally, makes a graphic. This routine is made to use, the "**SingleCellWithInterference**" scenario.

You should install the gnu plot software in your system to see the graphics.

$ sudo apt-get install gnuplot

1.5 Running the 5G air simulator

Open doSim1.sh and set all required parameters. In doSim1.sh, this path is shown. "./5G-air-simulator SingleCellWithI" this is the path I have used and it should be changed based on location of 5G-air-simulator file.

In all files search this "$./5G-air-simulator/TOOLS/./make-cell-spectral-efficiency$", this is the path for this research, it should be changed based on location of "TOOLS" file.

In the following files, “compute-throughput.sh”, “compute-spectral-efficiency.sh”, “fairnessIndex-comp.sh” ,“throughput-comp” and “spectral-efficiency-comp.sh” the "TIME" is set with 120 or 150, change this value for the time of your simulation.

Give permissions to all shells using:

$ chmod 777 shell-name.sh.

Finally run doSim1.sh.

BIBLIOGRAPHY

3GPP. (2018). System architecture for the 5G system. *3GPP TS 23.501 V15. 3.0*.

3GPP. (2022). Physical channels and modulation (3GPP TS 38.211 version 17.1.0 Release 17). *3GPP TS 138 211 V17.1.0 (2022-04)*.

Agyapong et al. (2014). Design considerations for a 5G network architecture. *IEEE Communications Magazine*, 52(11), 65–75.

AL-ALI, M. (2021). *Dynamic resource allocation of EMBB/URLLC traffic in 5G NR*. (Master's thesis).

AlQahtani et al. (2020). An efficient resource allocation to improve QoS of 5G slicing networks using general processor sharing-based scheduling algorithm. *International Journal of Communication Systems*, 33(4), e4250.

Alsenwi et al. (2019). A chance constrained based formulation for dynamic multiplexing of embb-urllc traffics in 5g new radio. *2019 International Conference on Information Networking (ICOIN)*, pp. 108–113.

Alsenwi, M. & Hong, C. S. (2018). Resource scheduling of URLLC/eMBB traffics in 5G new radio: A punctured scheduling approach. *Proc. Korean Comput. Congr.(KCC)*, pp. 1271–1273.

Benjebbour et al. (2018). 3GPP defined 5G requirements and evaluation conditions. *NTT DOCOMO Technical Journal*, 19(3), 13–23.

Campolo, G. et al. (2017). Better platooning control toward autonomous driving: An LTE device-to-device communications strategy that meets ultralow latency requirements. *IEEE Vehicular Technology Magazine*, 12(1), 30–38.

Chagdali et al. (2020). Impact of Slice Function Placement on the Performance of URLLC with Redundant Coverage. *2020 16th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, pp. 1–6.

Chamola et al. (2020). A comprehensive review of the COVID-19 pandemic and the role of IoT, drones, AI, blockchain, and 5G in managing its impact. *Ieee access*, 8, 90225–90265.

Chuah et al. (2002). Characterizing packet audio streams from internet multimedia applications. *2002 IEEE International Conference on Communications. Conference Proceedings. ICC 2002 (Cat. No. 02CH37333)*, 2, 1199–1203.

Dahlman et al. (2020). *5G NR: The next generation wireless access technology*. Academic Press.

Dutta, S. et al. (2019). A case for digital beamforming at mmWave. *IEEE Transactions on Wireless Communications*, 19(2), 756–770.

Fettweis et al. (2014). The tactile internet-itu-t technology watch report. *Int. Telecom. Union (ITU), Geneva*.

Ford, R. et al. (2017). Achieving ultra-low latency in 5G millimeter wave cellular networks. *IEEE Communications Magazine*, 55(3), 196–203.

Garcia-Perez et al. (2016). Enabling low latency services on LTE networks. *2016 IEEE 1st International Workshops on Foundations and Applications of Self* Systems (FAS* W)*, pp. 248–255.

Guan, P. et al. (2016). Ultra-low latency for 5g-a lab trial. *arXiv preprint arXiv:1610.04362*.

Guevara et al. (2020). The role of 5G technologies: Challenges in smart cities and intelligent transportation systems. *Sustainability*, 12(16), 6469.

Hänninen, N. (2020). Augmented reality in an amusement park environment: AR concept for Linnanmäki.

Hossain et al. (2021). 5G Multi-Band Numerology-Based TDD RAN Slicing for Throughput and Latency Sensitive Services. *IEEE Transactions on Mobile Computing*.

Hovila et al. (2019). 5g networks enabling new smart grid protection solutions.

Huseb et al. (2018). *Supporting Facilitators of Collaborative Learning using Mixed Reality-Helping Experts in Teamwork Facilitators observe groups collaborating in Virtual Reality*. (Master's thesis, NTNU).

Kasgari et al. (2019). Human-in-the-loop wireless communications: Machine learning and brain-aware resource management. *IEEE Transactions on Communications*, 67(11), 7727–7743.

Kayastha et al. (2014). Smart grid sensor data collection, communication, and networking: a tutorial. *Wireless communications and mobile computing*, 14(11), 1055–1087.

Lai, W. K. & Tang, C.-L. (2013). QoS-aware downlink packet scheduling for LTE networks. *Computer Networks*, 57(7), 1689–1698.

Lawal et al. (2017). Downlink scheduling algorithms in LTE Networks: A survey. *IOSR J Mob Comput Appl*, 4(3), 1–12.

Lee, H. & Ko, Y.-C. (2021). Physical Layer Enhancements for Ultra-Reliable Low-Latency Communications in 5G New Radio Systems. *IEEE Communications Standards Magazine*, 5(4), 112–122.

Lema, M. A. et al. (2017). Business case and technology analysis for 5G low latency applications. *IEEE Access*, 5, 5917–5935.

Li et al. (2018). 5G URLLC: Design challenges and system concepts. 1–6.

Madi et al. (2018). Delay-based and QoS-aware packet scheduling for RT and NRT multimedia services in LTE downlink systems. *EURASIP Journal on Wireless Communications and Networking*, 2018(1), 1–21.

Mamman et al. (2019). Quality of service class identifier (QCI) radio resource allocation algorithm for LTE downlink. *PloS one*, 14(1), e0210310.

Martiradonna, S., Grassi, A., Piro, G. & Boggia, G. (2020). Understanding the 5G-air-simulator: A tutorial on design criteria, technical components, and reference use cases. *Computer Networks*, 177, 107314.

Marzetta, T. L. (2010). Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas. *IEEE Transactions on Wireless Communications*, 9(11), 3590-3600.

Masum, M., Babu, M. et al. (2011). End-to-end delay performance evaluation for VoIP in the LTE network.

Mataj, E. (2020). *Network slicing and QoS in 5G systems and their impact on the MAC layer*. (Ph.D. thesis, Politecnico di Torino).

Meng et al. (2019). Integration application of 5g and smart grid. *2019 11th International Conference on Wireless Communications and Signal Processing (WCSP)*, pp. 1–7.

Navarro-Ortiz, J. et al. (2020). A survey on 5G usage scenarios and traffic models. *IEEE Communications Surveys & Tutorials*, 22(2), 905–929.

Osseiran et al. (2016). *5G mobile and wireless communications technology*. Cambridge University Press.

Pandey et al. (2019). A downlink resource scheduling strategy for URLLC traffic. *2019 IEEE international conference on big data and smart computing (BigComp)*, pp. 1–6.

Papadopoulos, H. et al. (2016). Massive MIMO technologies and challenges towards 5G. *IEICE Transactions on Communications*, 99(3), 602–621.

Parvez et al. (2018). A survey on low latency towards 5G: RAN, core network and caching solutions. *IEEE Communications Surveys & Tutorials*, 20(4), 3098–3130.

Pedersen et al. (2018). Agile 5G scheduler for improved E2E performance and flexibility for different network implementations. *IEEE Communications Magazine*, 56(3), 210–217.

Piro, G. et al. (2010). Simulating LTE cellular systems: An open-source framework. *IEEE transactions on vehicular technology*, 60(2), 498–513.

Pocovi et al. (2016). On the impact of multi-user traffic dynamics on low latency communications. *2016 International Symposium on Wireless Communication Systems (ISWCS)*, pp. 204–208.

Pocovi, G. et al. (2018). Achieving ultra-reliable low-latency communications: Challenges and envisioned system enhancements. *IEEE Network*, 32(2), 8–15.

Rao, J. et al. (2018). Packet duplication for URLLC in 5G: Architectural enhancements and performance analysis. *IEEE Network*, 32(2), 32–40.

Richart, M. et al. (2016). Resource slicing in virtual wireless networks: A survey. *IEEE Transactions on Network and Service Management*, 13(3), 462–476.

Rivas, G. et al. (2018). I choose you!: Gaming as a digital learning ecosystem for Early Childhood Education.

Roy et al. (2015). Location-based social video sharing over next generation cellular networks. *IEEE Communications Magazine*, 53(10), 136–143.

Salami et al. (1998). Design and description of CS-ACELP: A toll quality 8 kb/s speech coder. *IEEE transactions on Speech and Audio Processing*, 6(2), 116–130.

Sarwat, P. et al. (2018). Toward a smart city of interdependent critical infrastructure networks. In *Sustainable interdependent networks* (pp. 21–45). Springer.

Schulz,Riedel, I. et al. (2017). Latency critical IoT applications in 5G: Perspective on the design of radio interface and network architecture. *IEEE Communications Magazine*, 55(2), 70–78.

Seeling et al. (2011). Video transport evaluation with H. 264 video traces. *IEEE Communications Surveys & Tutorials*, 14(4), 1142–1165.

Series, M. (2015). IMT Vision–Framework and overall objectives of the future development of IMT for 2020 and beyond. *Recommendation ITU*, 2083, 0.

Shakkottai et al. (2001). Scheduling algorithms for a mixture of real-time and non-real-time data in HDR. In *Teletraffic Science and Engineering* (vol. 4, pp. 793–804). Elsevier.

Siddiqi, M. A. et al. (2019). 5G ultra-reliable low-latency communication implementation challenges and operational issues with IoT devices. *Electronics*, 8(9), 981.

Simsek et al. (2016). 5G-enabled tactile internet. *IEEE Journal on Selected Areas in Communications*, 34(3), 460–473.

Sutton, G. et al. (2019). Enabling technologies for ultra-reliable and low latency communications: From PHY and MAC layer perspectives. *IEEE Communications Surveys & Tutorials*, 21(3), 2488–2524.

Sybis, M. et al. (2016). Channel coding for ultra-reliable low-latency communication in 5G systems. *2016 IEEE 84th vehicular technology conference (VTC-Fall)*, pp. 1–5.

Wang, W. et al. (2017). Field trial on TDD massive MIMO system with polar code. *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pp. 1–6.

Wilner et al. (1976). Fiber-optic delay lines for microwave signal processing. *Proceedings of the IEEE*, 64(5), 805–807.

Wu et al. (2015). Millimeter-wave multimedia communications: challenges, methodology, and applications. *IEEE communications Magazine*, 53(1), 232–238.

Zaidi et al. (2018). *5G Physical Layer: principles, models and technology components*. Academic Press.

Zhan et al. (2021). Performance Optimization for Massive Random Access of mMTC in Cellular Networks With Preamble Retransmission Limit. *IEEE Transactions on Vehicular Technology*, 70(9), 8854-8867. doi: 10.1109/TVT.2021.3096259.