Tampere University

Sarthak Seth

CONTENTION BASED RESOURCE ALLOCATION IN 6G

Master of Science Thesis
Faculty of Information Technology and Communication Sciences
Examiner: Prof. Mikko Valkama
Dr. Jukka Talvitie
May 2022

ABSTRACT

Sarthak Seth: Contention based resource allocation in 6G
Master of Science Thesis
Tampere University
May 2022

Reliable and Low Latency Communications is one of the several use cases targeted by 5G NR. Considering this, 5G has standardized a Grant-Free transmission method in the form of Configured Grant for dedicated and reliable low latency access. Configured Grant is highly suitable for deterministic traffic but can be a spectrally inefficient for sporadically transmitting users. On the other hand, contention or shared pool-based approaches could provide spectrally efficient grant-free access for group of sporadically transmitting users. However, users can select the same resource in the pool, and collisions can happen causing a degradation in reliability. In this thesis work, we aim to study a useful sensing design for users sending over a grant-free channel how future 6th generation networks will enable shared resources framework for allocating resources. The sensory process can help in assisting users in eliminating collisions and cutting delay.

Preliminary results have shown, for certain traffic type, the sensing methods can outperform existing scheduling methods, namely dynamic and configured grant in terms of performance. This is supported by my mathematical and graphical results and thus can be considered a viable candidate for scheduling methods expansion in 6G to serve versatile traffic.

Keywords: 5G, NR, 6G, Grant-Free, contention, sensing, uplink, low latency, Configured Grant

PREFACE

This thesis is based on studies, modeling, and research done at Ericsson, Jorvas and is written as a finalization of the master's degree education in Information Technology and Communication Sciences in Tampere University. I appreciate my supervisors' advice and constructive criticism during the thesis, which allowed me to gain sufficient knowledge regarding fifth-generation mobile networks and beyond and I was able to publish a conference paper based on my thesis work. I mastered several research procedures and scientific writing techniques while working on my thesis. Finally, I would like to express my gratitude to my family and friends for their unwavering support and love for me.

Tampereella, 18th May 2022

Sarthak Seth

CONTENTS

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

1G	1st Generation
2G	2nd Generation
5G	5th Generation
6G	6th Generation
AWGN	Additive White Gaussian Noise
BER	Bit Error Rate
BPSK	Binary Phase Shift Keying
CG	Configured Grant
eMBB	enhanced Mobile Broadband
FDD	Frequency Division Duplex
GF	Grant-Free
hspa	High Speed Packet Access
MIMO	Multiple Input Multiple Output
mMTC	massive Machine-Type Communications
NR	New Radio
OFDM	Orthogonal Frequency Division Duplex
PRB	Physical Resource Block
PUSCH	Physical Uplink Shared Channel
RACH	Random Access channel
SNR	Signal-to-Noise Ratio
TAU	Tampere University
TB	Transport Block

TDD	Time Division Duplex
TUNI	Tampere Universities
UE	User Equipment
URL	Uniform Resource Locator
URLLC	Ultra-Reliable and Low Latency Communication
ZF	Zero Forcing

1. INTRODUCTION

Ultra-Reliable and Low Latency Communications (URLLC), are one of the most important use cases in the 5th Generation (5G) New Radio (NR) standard, as defined by 3GPP Release 15 (3rd Generation Partnership Project). URLLC will support a lot of sophisticated latency-sensitive linked devices in different fields such as vehicle technology, artificial intelligence, virtual reality and many more. Other important use case in 5G is eMBB (Enhanced Mobile Broadband) which will provide high-bandwidth internet access for wireless networking, large-scale video streaming, and virtual reality. And mMTC (Massive Machine Type Communication), which allows sensing, metering, and monitoring devices to connect to the internet. This thesis focuses on scenarios involving URLLC services that require reliability of 99.999 percent and with latency down to 0.5 to 1 millisecond [1]. The LL, or low latency, is one of URLLC's most important features. Low latency is critical for devices that drive them-selves or perform prostate operations, for example. Low latency allows a network to be tailored for processing massive amounts of data with the least amount of. In real time, the networks must react to a large amount of changing data. URLLC is perhaps the most promising component of impending 5G capabilities [2,3].

In long-term evolution (LTE) as well as earlier in 5G grant-based (GB) scheduling was used. It involved several handshakes between base station and user equipment (UE) in form of scheduling request (SR) and scheduling grant (SG). This caused a lot of delay which is not suitable for 5G NR and future generations. To enable low latency access grant-free (GF) scheduling was targeted. Here the resources are preassigned to users which eliminates the delay caused in GB scheduling. Configured Grant (CG) was introduced in NR [4], it supports a periodic traffic to readily transmit its data without any delay caused due to scheduling request and scheduling grant [5]. But for sporadic traffic CG is not a suitable choice as there may be instances where the resources are not used hence resulting in a resource inefficient method [6].

The employment of shared resource by a group of UEs better option which could minimize the resource wastage but in CG we need a dedicated downlink control information (DCI) for each user thus require treating every UE individually puts more burden on control resource utilization. In a contention pool as in [7] every UE does not need to be treat-ed individually and resource utilization is also improved. But there can be collision with the contention which can lead to loss in reliability and bit error rate. The current GF scheduling can be classified into two types of schemes: reactive and proactive. Reactive transmission is a mechanism in which the UE transmits a packet again if the previous one is not received, improving resource consumption. The latency of the system is affected by the feedback procedure used in this method. Another option is proactive transmission, which repeats the transmission until the UE delivers favourable feedback, however this is harder for the UE to calculate because it must constantly check for an acknowledgement [8,9].

K repetition is a proactive method in which some K repeats of the same packet are delivered in consecutive slots to ensure reliable transmission; it enhances redundancy if the preceding repetition was successfully received, but it wastes resources [10]. It may also unnecessarily raise the burden and entail statistical complexity. To improve the results of these techniques, various methods have been offered. Adaptive repetition is used in [11] to reduce the likelihood of a collision. The fading cor-relation, which is utilized to limit the number of repetitions, is derived through sensing. The adoption of improved receivers could prevent collisions, according to [11], but the complexity involved increases power and latency. In [12], a new sensing strategy is described that uses priority-based contention and announcement messages. This method outperformed earlier methods in terms of resource consumption for intermittent traffic. The issue here is that announcement message transmission is prioritized, which limits the maximum number of UEs. Furthermore, the priority-based strategy is unjust to peer nodes. The system is complicated by the priority-based announcement messages, and the sensing is based on minislots, which wastes resources.

Contention-based allocation enables many UEs to simultaneously attempt to send their messages over the same channel or resource. The resources can be set up so that UEs contend or compete for their transmissions in a contention pool. If the same resource is selected and no collision avoidance techniques are applied, there may be a collision. Loss of dependability or bit rate results from this. We have presented two sensing-based methods that are described in comparison to standard techniques that involve blind repeats in order to prevent collisions and reliability loss. It is crucial to note that actual contention occurs when UEs can avoid collisions thanks to implemented policies and features.

We have developed a new technique to full duplex operation using a sensing mechanism in this thesis. The UE transmits an orthogonal sequence, such as Demodulation Reference Signal (DMRS), which is already included in 5G NR. If the UE hears other UEs sending their sequence, it backs off and performs a conditional retransmission of the packet. This improves reliability without significantly increasing delay. In addition, the second proposal in this work is to use a conditional grant method in which the packet is delivered in a dedicated grant channel if more than one UE is transmitting, making the system collision-free and therefore enhancing system performance. An error rate comparison has been performed among the primitive contention-based methodologies to analyse and evaluate the collision rate. This increases reliability without significantly raising latency. It is possible to use the sensor mechanism in full- or half-duplex mode. The error rates of the suggested schemes are compared to those of multiple K repetitions schemes, and gains are then shown as a result. Finally, a conclusion is reached that considers the likelihood of such plans being implemented in 6G networks.The work done in this thesis has also been published[19].

The rest of this thesis is organized as follows. Chapter 2 gives a general overview of prevailing scheduling schemes, discusses about them and try to explore more about the existing problems and why they are not suitable in the future generations of communication networking. Chapter 3 focuses on contention, as in what is contention, why it is used and how it is advantageous for future generations. In

Chapter 4 and chapter 5 new model is introduced with proper mathematical equations and comparison between existing methods and the proposed scheme with the help of equations and discussions is also done. Finally, simulation results are used to compare the existing and new models discussed in the thesis and proper reasoning is given why the new schemes are better.

2. FIFTH-GENERATION NEW RADIO

2.1 Motivation

There has been a revolution in mobile communication over the previous four decades. From the First Generation (1G) to the Fifth Generation (5G), these improvements have been gradually implemented. Around 1980, the Nordic Mobile Telephony (NMT) system, the Advanced Mobile Phone System (AMPS) in North America, and the Total Access Communication System (TACS) in the United Kingdom developed 1G mobile communication technology. The core was an analog communication network that provided voice services to a small number of users.

The Second Generation (2G) of mobile communication, also known as the Global System for Mobile Communication (GSM), was launched in the early 1990s. Digital AMPS (D-AMPS) and Personal Digital Cellular (PDC) helped to the development of 2G in Japan, and IS-95/CDMA and IS-136/TDMA technologies were employed in America [14]. This generation saw the introduction of new digital transmission between radio connections, which improved the quality of voice calls but limiting data services. The technology gradually extended from Europe to all the parts of the globe, and mobile telecommunications companies began to employ it extensively.

The Third Generation (3G) of mobile communication was introduced in early stages of 2000s. The key advancement in this generation was the availability of high-speed mobile broadband. High Speed Packet Access (HSPA), a later version of 3G provided high speed connection, which is still used in some parts of the world. [15]. The unpaired spectrum was introduced with 3G, which was based on Time Division Synchronous Code Division Multiple Access (TD-SCDMA), which utilised Time Division Duplex (TDD). Previously, mobile communication was meant to operate in paired spectrum using Frequency-Division Duplex technology (FDD).

In 2009, the initial technical requirements for fourth-generation mobile communications, often known as LTE, were released. It follows in the footsteps of HSPA by adopting an OFDM (Orthogonal Frequency Division Multiplexing)-based transmission system to provide end-users with better data rates and increased efficiency. It also makes use of a broader spectrum and innovative multi-antenna systems. Within one common radio-access technology, FDD and TDD are both supported in 4G, which are operating in paired or maybe unpaired frequencies. In addition to standard broadband data connections, LTE provides a range of supplemental use cases, such as minimal gadgets with good battery life, mMTC services, and lower air-interface latency. As a result, LTE's evolution is capable of supporting a variety of 5G use cases.

The LTE envisage technologies have been remodeled thanks to NR technology. The evolution of LTE and NR are inextricably linked. Because LTE has been in use for more than a decade, newer, more rigorous technological standards are unlikely to be met. As a result, in the fall of 2015, the 3GPP convened a workshop to satisfy the new requirements and specifications for advanced technologies. As a result, new radio-access technology, today known as NR, was developed. The first edition of the NR standard was issued at the end of 2017, and the first 5G design was commercially implemented in 2018.

The term 5G refers to the 5G of all-digital wireless cellular technology. Although we normally refer to 5G NR, which is the 3GPP-standardized new radio access technology, the term 5G is often used in a broader sense (RAT). 5G Standards are approved by the International Telecommunication Union's Radio communication Sector (ITU-R). The International Telecommunication Union's International Mobile Telecommunication IMT-2020 is noted for its 5G standards, which include pre-commercial operations and 5G trials to aid in evaluating potential technologies and frequency bands.

Parallel to the ITU-R schedule, the 3GPP follows the standardization procedure. In mid-2017, the 3GPP technical specifications groups agreed on a detailed work schedule for Publication 15, resulting in the first release of 5G specifications. Approximately one 3GPP publication is released each year. These publications look into both RAN and the essential components of NR.The 38-series for NR defines the radio part of 3GPP. The first specification in Release 15 is for non-standalone NR, which relies on 4G for mobility and initial connectivity.

2.2 Use Cases

The 5G triangle or the three major areas to be benefited from 5G are:

below.

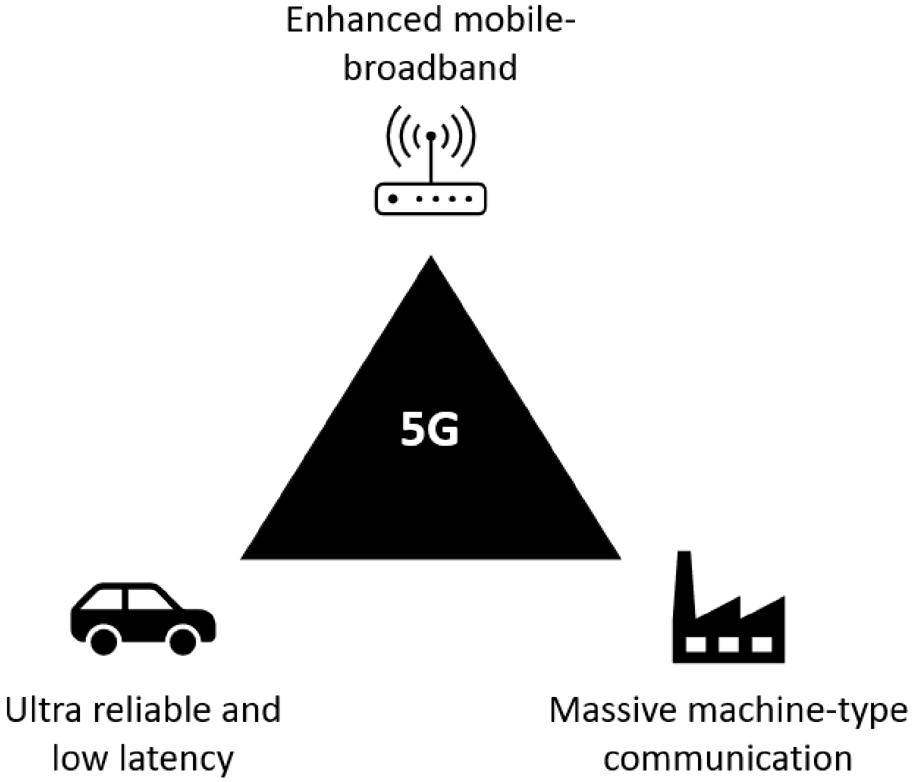


Figure 2.1. *Key 5G use cases.*

2.2.1 eMBB:Enhanced Mobile Broadband

5G provides extraordinarily high data rates in order to support high user density and capacity. Human-centric communication and reliable connection over vast coverage areas will be the focus of Enhanced Mobile Broadband. Users will be able to enjoy the high speed internet features such as AI and VR with high definition video and picture quality.

For example, businesses will be able to store more data in the cloud and access it as if it were stored locally over fast, low-latency 5G. This eliminates the need for costly on-site servers. Users will not need high tech systems to fetch the information or data instead user can fetch data from the cloud within a second. The experience will be like simply having the file saved locally.

2.2.2 mMTC:Massive Machine Type Communication

Machine-type communication will necessitate the connection of millions of network-enabled objects, such as in the context of the Internet of Things (IOT). IoT's main challenges will be the expansion of technology network. It is believed that, the number of IoT linked devices per individual on the earth will rise from 2 per person currently to 10 per person by 2025. The anticipated network of gadgets that require a data connection puts enormous pressure on network infrastructure, such as mobile communication towers. Introduction to 5G will improve this network and help in expansion of connected devices as it is in 4G.

2.2.3 uRLLC:Ultra Reliable Low Latency Communication

Once 5G is fully deployed, ultra-reliable low-latency communication (uRLLC) will be one of the most significant change.self driving vehicles are one of the most important application of 5G communication. The technology related to vehicle industry requires a high speed and high reliability. Computer systems have progressed that they can now equal data center processing capabilities.

A vehicle-to-everything (V2X) communication network is the what we want to have in our future. A self changing mechanism will be the future of vehicle industry where they can change or respond to the changes that occur in daily basis such as climate and weather. A vehicle must be able to communicate and receive data in a few milliseconds in order to stop or adjust directions in response to signs, threats, and individuals on the road.

Bidirectional communication between vehicle and infrastructure will be enabled by connected car technologies, mitigating risk across transportation networks. In cities, sensors are now being deployed in every junction to find changes and trigger connected and technology and vehicles to make changes as needed. The communications backbone for linked car technology might be installed, enhancing pedestrian and vehicle safety dramatically.

AR and VR applications will be more improved and full of new features with 5G's low latency. A technician wearing AR wearable, may help to do things that could not be done with bare eyes, providing different instructions or maybe help to identify parts, or display things that are dangerous to touch in an industry.

2.3 Features

2.3.1 Capacity

OFDM (Orthogonal frequency-division multiplexing) is a technology for decreasing interference by modulating a digital transmission across many channels, which is used in 5G. A 5G NR air interface is used in conjunction with OFDM principles in 5G. 5G also employs higher-bandwidth technologies such as sub-6 GHz and mmWave. The concepts used in 5G are similar to as that of 4G OFDM. But the capacity to handle users increase in 5G drastically over 4G. The need to connect more and more devices in the future can be fulfilled by 5G communication network and hence it is an important important feature.

The spectrum is stretched in 5G to several hundreds of GHz which will make more bandwidth available for users to use hence increasing the overall capacity of the network.

2.3.2 Massive Connectivity

5G is set to bring in a future that is completely connected. Thanks to the tremendous increase in data speeds and the number of linked gadgets in 5G network user experience will improve. By 2025, 5G is planned to link a staggering 25 billion devices. 5G is not just a transformation to mobile networks but also a strategy to combine mobile networks with wifi, and can be thought of as an overlay over the existing 4G network.

5G will improve the user experience in literally all the fields where one can imagine. 5G network will bring massive connectivity between devices of different fields whether it is vehicle technology, healthcare of transforming a city into smart city.

2.3.3 Low Latency

Latency is the amount of time that passes between a user's request and its completion in the IT world. Even seemingly instantaneous processes have a quantifiable lag. Reducing such delays has become a top priority for any organization.

The most crucial feature of 5G technology, according to popular belief, is faster internet speeds. Many people overlook the fact that 5G can be used to address a far more pressing issue: network latency reduction. The end-to-end communication delay is defined as the time between delivering a piece of data and receiving the appropriate response.

The latency rate, or time between sending and receiving data, is incredibly low with 5G technology. With 5G, we can go from 200 ms to 1ms. Low latency enabled by the introduction of 5G networks opens the door to completely new perspectives, such as real time gaming, virtual reality , advancement in robotics, self-driving cars, and many more where a quick reaction is not a choice, but a necessity.

2.3.4 High Reliability

The new era of communication is always believed to be the high speed networks and unlimited data with massive device support, but one of the most important and necessary feature is reliability. If a connection or network is not sound enough then the high speed is of no use.

The autonomous driving and other application which require crucial data transfer reliability is important.Connected vehicle technology will enable bidirectional communications between vehicle and infrastructure to increase safety. In cities, sensors are now being deployed to track the changes and trigger the connected vehicles to make changes accordingly. The network backbone for linked car technology might be installed now, enhancing pedestrian and vehicle safety dramatically.

3. SIXTH GENERATION LOW LATENCY COMMUNICATION

3.1 Motivation

Wireless technologies are useful today because they allow users to wirelessly transfer data from one base station to another. These innovations enable us to communicate with users over short and large distances. Users can converse instantaneously with the use of wireless communication. Many wireless technologies, such as 1G-5G, are now available, however they do not provide high availability, network performance, or coverage that could be enough for a technology with very high rate internet with minimum latency. 6G will allow devices to connect to the internet via wifi, wimax, or maybe Bluetooth.

Mobile network technologies such as 5G are now available on practically all mobile handsets. However, network data speed and coverage continue to be a problem. With a data speed of around 10-11 Gbps, 6G has the potential to provide users with more than they expect. Currently, technologies provide Internet speeds of up to 100-500 Mbps. Additionally, 6G capability is available to improve transmission of data and wireless security. The 6G design is primarily utilized to achieve worldwide network coverage by combining the 5G with the satellite network. Satellite networks and telecommunication networks remain very useful for mobile network users and 6G plans on combining both of them by connecting a base station to the satellite. In this way a whole lot area could be covered and with very high data rate. The research on 6G is still in progress and there are different concepts in different papers defining 6G.

The THz band is the key hurdle in the 6G communication technology. Despite the high data speeds, the high frequencies make tackling the high path loss a significant challenge. The air absorption and transmission loss for long-distance communications are extremely high. This is a critical problem that must be addressed. To tackle the issue of frequency dispersion, new multipath channel models must be devised.

3.2 Shortcomings of Existing Technologies

3.2.1 NR CG

The URLLC service was first supported in 3GPP Release 15. The URLLC service was substantially enhanced in Release 16 with features such as configured grant (CG) based transmission. The CG based uplink (UL) transmission system can drastically minimize transmission latency by omitting the scheduling request (SR) and uplink grant procedures from gNodeB (gNB), and a user equipment (UE) can trigger an uplink transmission once the data packet arrives. When a URLLC packet arrives at the UE, it will choose the latest available configured resource for uplink transmission, and the corresponding CG timer will be triggered in the meantime. Then, based on the gNB's behavior, there will be two scenarios to consider. Before the timer ends, the UE receives an uplink grant from the gNB for packet re-transmission. The UE then retransmits the packet based on the obtained grant. Note that demodulation reference signal (DMRS) detection is only required for CG-based PUSCH transmissions; it is not required for dynamic grant (DG) based PUSCH transmissions because the gNB already knows what it is supposed to do. The UE receives nothing until the timer expires, at which point it assumes that the gNB correctly decoded the TB, implying that the UE assumes the implicit ACK in this situation.

However, CG is not a good choice for intermittent traffic because there may be times when the resources are not utilised, resulting in a resource wasteful method. The usage of a shared resource by a group of UEs is a preferable alternative for reducing resource waste, but in CG, we need a distinct downlink control information (DCI) for each user, which necessitates handling each UE separately, which increases the strain on control resource utilization. Every UE does not need to be treated individually in a contention pool, and resource usage is also improved [17].

3.2.2 NR 2-Step RACH

There was only one sort of RACH procedure defined in earlier versions of NR 3GPP Releases, such as Rel.15, which is comparable to LTE RACH method, also known as 4-Step RACH procedure. 2-Step RACH technique was introduced in Release-16 NR, which attempts to reduce the total latency of the RACH procedure. The advantage of a 2-Step RACH process over a 4-Step RACH method is that it lowers network access latency. The base station and the UE engage twice in a typical 4-step Random Access Channel (RACH). Msg1 is used to determine timing advance, and Msg3 is used for UE identification. This approach is reliable, but it is inefficient because it requires two interactions. Msg1 and Msg3 are combined in a single Physical Uplink Shared Channel (PUSCH) as MsgA, which is more efficient because it reduces latency and transmission power. In 2-step RACH, the lack of prior information about timing and resource allocation causes a contention-based asynchronous PUSCH. Introduction to contention causes collision and which ultimately affects the reliability of the system [16].

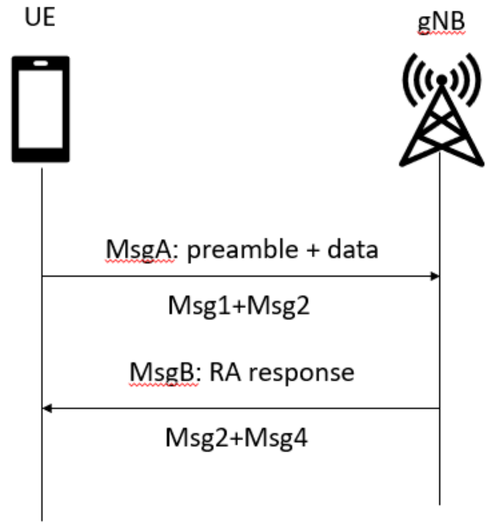


Figure 3.1. *Process involved in 2-step RACH.*

3.2.3 CSMA or LBT based mechanism

The 5G NR in unlicensed spectrum (NR-U) mode of operation is a new NR Release 16 mode of operation that gives cellular operators the technology they need to completely integrate unlicensed spectrum into 5G networks. NR-U supports 5G novel features like as wideband carriers, flexible numerologies, and dynamic scheduling/HARQ timing in unlicensed bands, enabling both uplink and downlink operation.

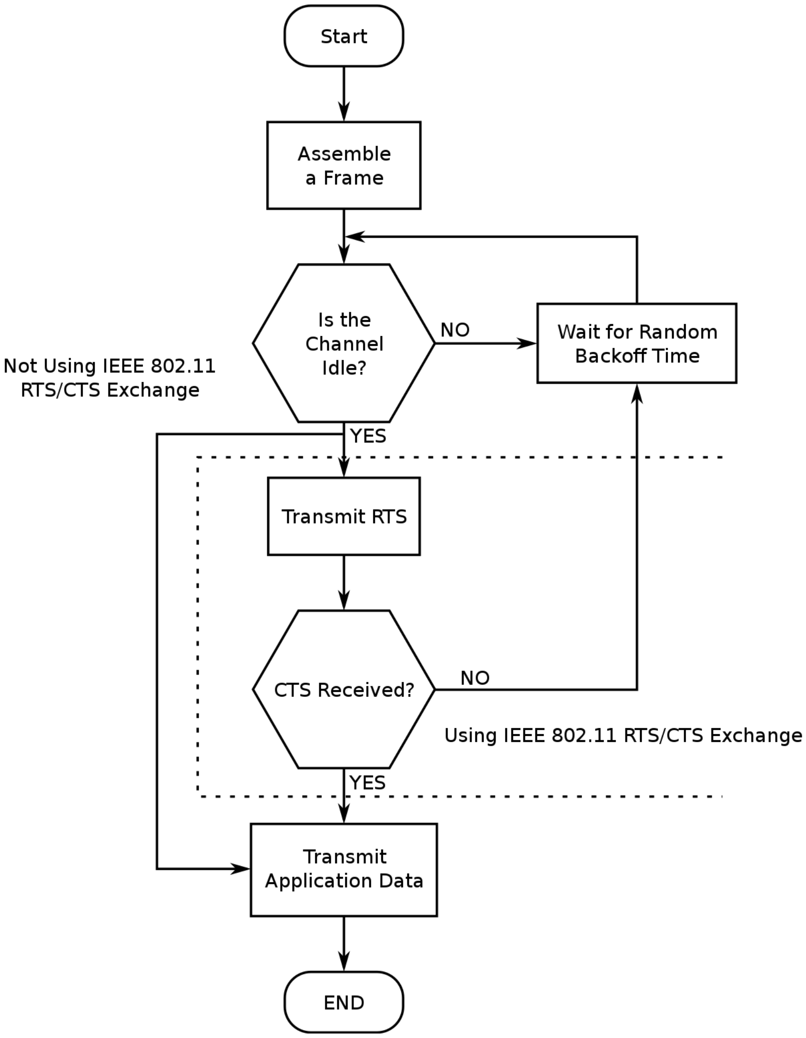


Figure 3.2. *Representation of CSMA or LBT mechanism.*

The listen-before-talk (LBT) feature is used in NR-U for both downlink and uplink channel access. Prior to transmitting, a wireless device or a base station must first "sense" the communications channel to ensure that no communications are present. The "channel sensing" approach relies on detecting the energy level on various sub-bands of the communications channel when the communication channel is a large bandwidth unlicensed carrier (e.g., several hundreds MHz). The base station configures the LBT parameters in a wireless device.

The 3rd Generation Partnership Project (3GPP) has created a new frame structure type 3 for use in unlicensed communications channels. Frame structure type 3 uses the same frequency channel as LTE time division duplex (TDD), but the uplink and downlink operations are separated in time. A subframe is not configured as a downlink or uplink subframe in LTE TDD, and it can be used by either the base station or the wireless device. They do, however, have the disadvantage of having to be known by neighboring devices, and they are only observable for the portion of the transmission where known format signals are present.

4. SYSTEM MODEL

4.1 Resource Pool and Issues

Contention occurs when two or more UEs attempt to send a message over the same channel or spectrum resource at the same time. A contention pool is a collection of resources for which the UEs contend or compete. If the same resource is picked, a collision occurs, resulting in a loss of reliability or bit rate, which is inconvenient for URLLC because the time budget is limited. There are numerous solutions to avoid this problem, which are detailed below.

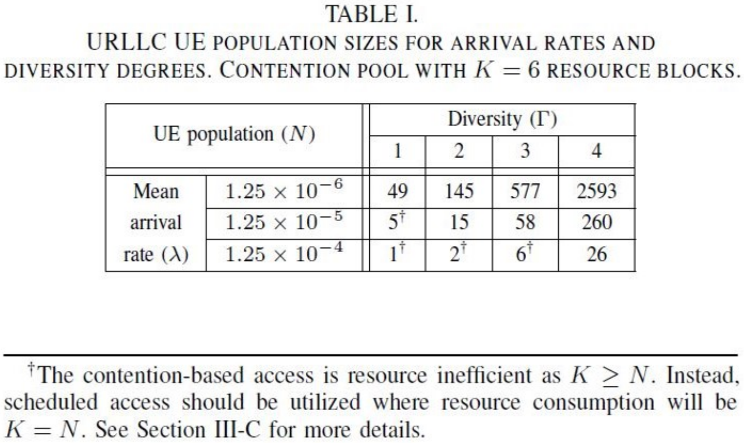


TABLE I.
URLLC UE POPULATION SIZES FOR ARRIVAL RATES AND DIVERSITY DEGREES. CONTENTION POOL WITH $K = 6$ RESOURCE BLOCKS.

UE population (N)		Diversity (Γ)			
		1	2	3	4
Mean arrival rate (λ)	1.25×10^{-6}	49	145	577	2593
	1.25×10^{-5}	$5^{\dagger}$	15	58	260
	1.25×10^{-4}	$1^{\dagger}$	$2^{\dagger}$	$6^{\dagger}$	26

---

†The contention-based access is resource inefficient as $K \geq N$. Instead, scheduled access should be utilized where resource consumption will be $K = N$. See Section III-C for more details.

Figure 4.1. *An example of Contention based resource pool.*

The frequency resource in GF transmission might be reserved in advance or allocated when a request is received. Preallocation of the dedicated resource, also known as Semi-Persistent-Scheduling (SPS), is better for fixed-pattern traffic, but contention-based GF transmission over the shared resource is better for sporadic packets since it is more resource-efficient and adaptable.

Contention-based GF transmission, on the other hand, is vulnerable to collisions with other UEs transmitting over the same resource at the same time, endangering transmission reliability.

4.2 Transmission Parameters

We consider a system where a load is generated during the access window using a discrete random variable X based on Poisson distribution with mean load $\lambda > 0$,

$$\mathcal{P}(\mathbb{X} = \ltimes) = \frac{e^{-\lambda}\lambda^n}{n!}.$$

where in above equation $\mathcal{P}$ is the probability when variable $X = n$ UE.The W slots and R PRB are selected at random by the UE. Each UE requires one PRB in a slot for its packet and selects one at random from a pool of available slots and PRB. There are W slots available, with R PRB per slot.

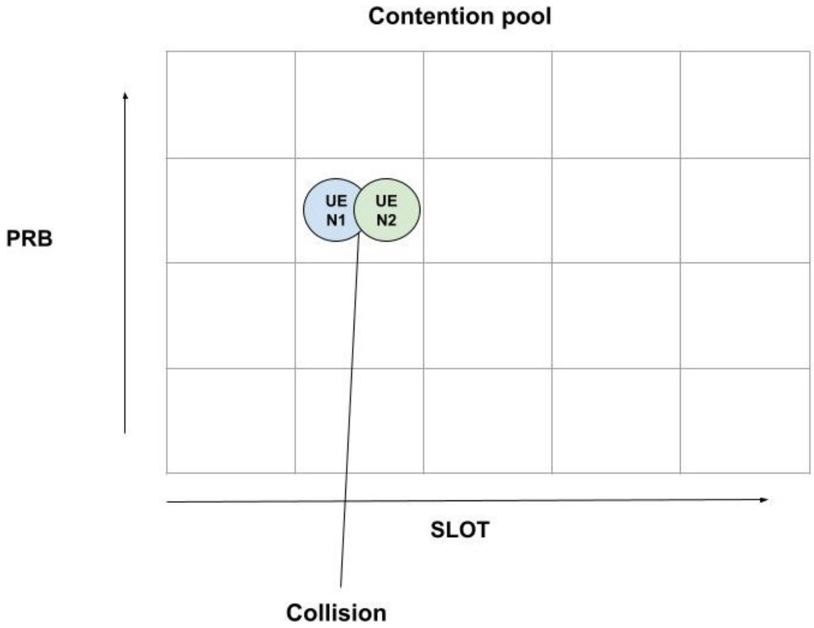


Figure 4.2. *Representation of a contention pool and collision.*

The packet is then sent across an MIMO channel. The data is produced at random and transferred to the appropriate symbol using ZF modulation. The channel also employs an ZF detector to counteract the impacts of the channel.

Rayleigh fading channel is the model type defined in this study. The ZF detector uses an inverse function to cancel out the channel's effects. The outcome of applying this inverse function on the chosen symbol is 1, whereas the result for all other symbols is 0. It is a successful transmission if the packet does not collide and is received without any errors.

4.3 Baseline

Contention-based allocation enables many UEs to simultaneously attempt to send their messages over the same channel or resource. The resources can be set up so that UEs contest or compete for their transmissions in a contention pool. If the same resource is selected and no collision avoidance techniques are applied, there may be a collision. Loss of dependability or bit rate results from this. We have presented two sensing-based approaches that are described in comparison to standard techniques that need blind repeats in order to prevent collisions and reliability loss. It is crucial to note that actual contention occurs when UEs can avoid collisions thanks to implemented policies and features.

4.3.1 No repetition

No repetition is a fundamental approach in which the UE merely transmits their packets by selecting an PRB from the pool, with no real resource conflict. There is collision and data loss if two UE choose the same PRB; otherwise, data is transferred via the channel and the error probability is determined. No collisions must occur in order to determine the success rate for an UE, and channel limitations should not be a source of error.

The probability analysis for BER is done to compare this model with the extended schemes discussed later in the paper.

From Poisson, we get the probability to have n UE transmitting as,

$$\mathbb{P}_{\mathrm{n}} = \frac{e^{-\lambda}\lambda^{n}}{n!}.$$

Then the probability that there is no UE transmitting is

$$\mathbb{P}_0 = e^{-\lambda}. \tag{4.1}$$

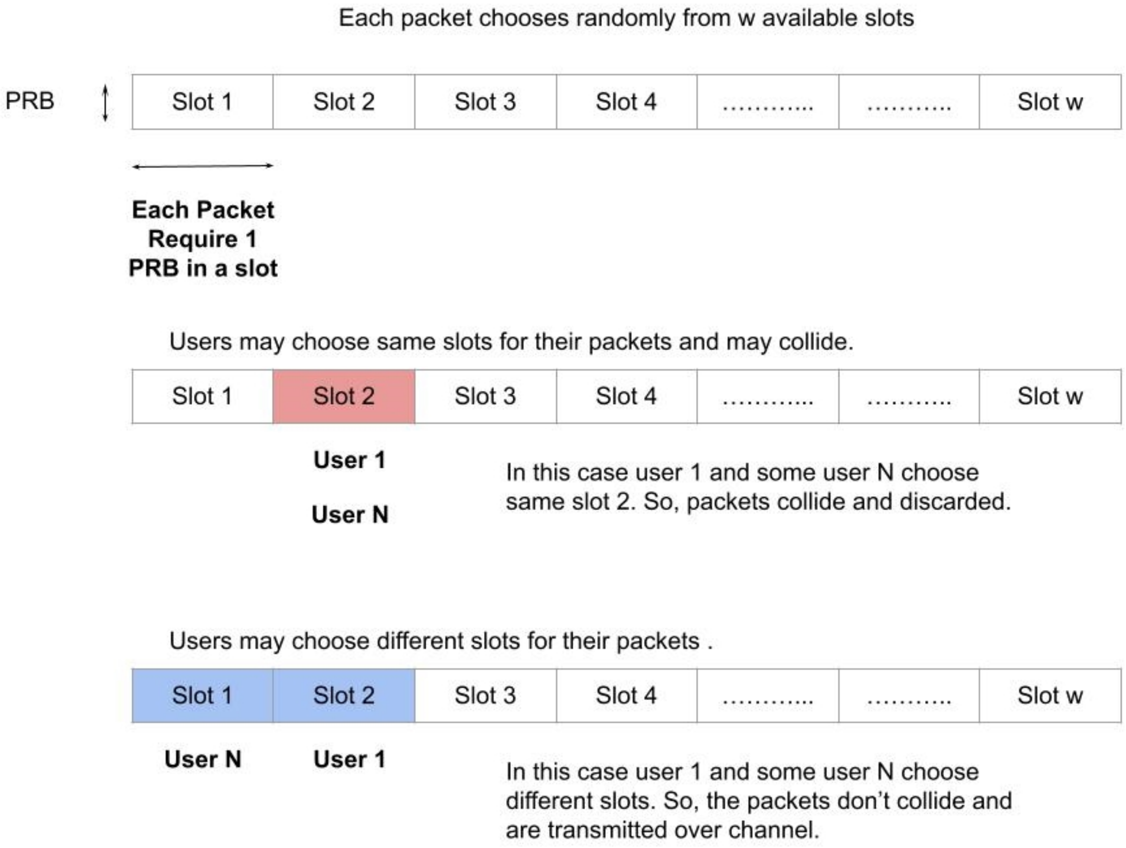


Figure 4.3. *The process involved in a no repetition method .*

The probability that there are n UE transmitting and they do not collide with each other is

$$\mathbb{P}_{-\mathrm{c}} = \sum_{n=1}^{\infty} \mathbb{P}_{\mathrm{n}} \left(\frac{WR-1}{WR} \right)^{n-1}. \tag{4.2}$$

Then the total collision probability from (4.1) and (5.5) is

$$\mathbb{P}_{\mathrm{c}} = 1 - \mathbb{P}_0 - \mathbb{P}_{-\mathrm{c}}. \tag{4.3}$$

For successful transmission of a packet there must be no collision and no channel impairments and the channel impairment probability is

$$\mathbb{P}_{\mathrm{BER}} = 1 - \sqrt{\frac{\mathrm{SNR}}{\mathrm{SNR}+1}}. \tag{4.4}$$

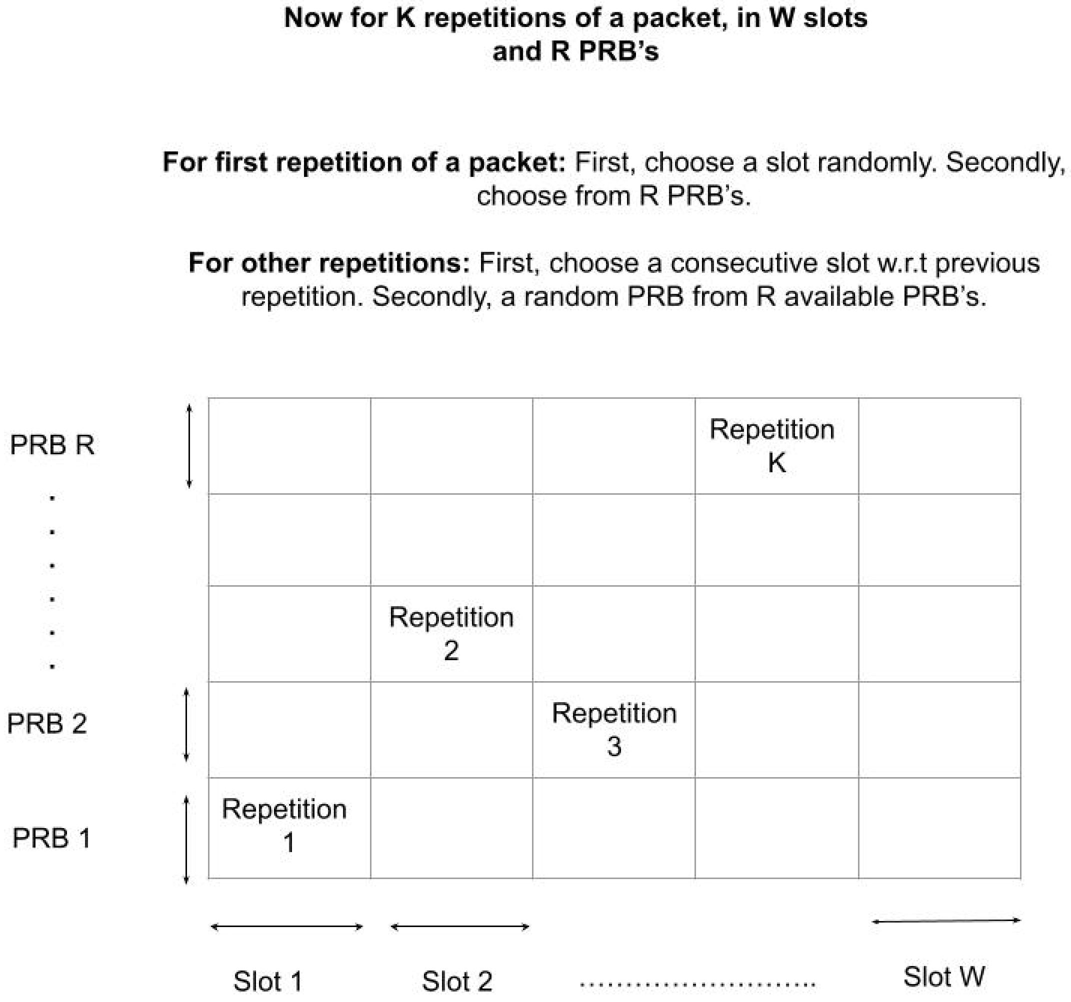
Then the probability of successful transmission is

$$\mathbb{P}_{\mathrm{s}} = (1 - \mathbb{P}_{\mathrm{c}})(1 - \mathbb{P}_{\mathrm{BER}}). \tag{4.5}$$

Then the failure probability of the system is

$$\mathbb{P}_{\mathrm{f}} = 1 - \mathbb{P}_{\mathrm{s}}. \tag{4.6}$$

4.3.2 Multiple repetition



Now for K repetitions of a packet, in W slots and R PRB's

For first repetition of a packet: First, choose a slot randomly. Secondly, choose from R PRB's.

For other repetitions: First, choose a consecutive slot w.r.t previous repetition. Secondly, a random PRB from R available PRB's.

Figure 4.4. *Representation of K repetition process and steps involved.*

To increase the rate of success, the same packet is repeated four times in this manner. A UE selects a random slot from W available slots for the first repeat of the packet, and then selects from available R PRB. For subsequent repeats, the UE selects a following slot and then selects a random PRB from among the available PRB. Only when all of the packets collide or become erroneous does a total packet loss occur. This minimizes the likelihood of an error, but at the cost of increased latency. We can get an analytical formula of collision probability for upper and lower bounds using.

The upper bound of collision probability is calculated by assuming that the K packets of a UE arrive in a synchronized manner that is in consecutive slots. The lower bound is calculated by assuming that the K packets of a UE does not arrive in consecutive slots and are fully independent.

The failure probability for multiple repetition from Eq. (4.6) is

$$\mathbb{P}_{\mathrm{f}}^{K\mathrm{rep}} = (\mathbb{P}_{\mathrm{f}}(K\lambda))^{K}.$$

5. RESOURCE ALLOCATION STRATEGIES

The URLLC standards are difficult to meet with the current Long Term Evolution (LTE) infrastructure. Current LTE uses a schedule-based transmission mechanism, particularly grant-based (GB) scheduling, especially in the uplink. The User Equipment (UE) initiates this traditional GB scheduling by sending an access request to the network, the BS can reply by providing access grant via a four-step random access mechanism. Such a scheduling-request-triggered transfer takes upto 10 milliseconds before beginning the transmission, greatly exceeding the URLLC latency threshold. To deal with this situation, grant-free (GF) access has recently been proposed and intensively studied.

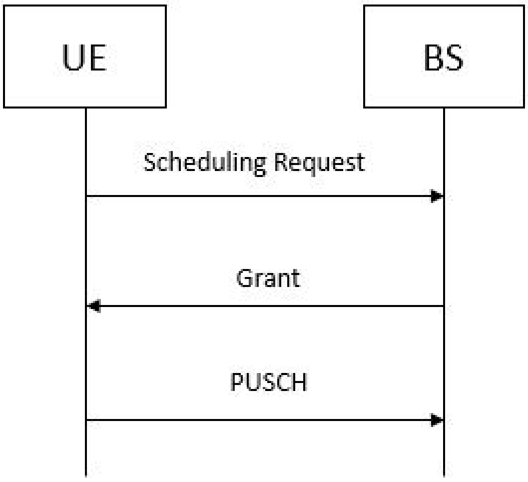


Figure 5.1. *Process in involved in grant-based scheduling.*

Grant-free (GF) random access is offered as a way to reduce access latency by skipping the handshake-based grant acquisition phase [4]. When using GF transmissions, a user with available traffic sends the data (together with the necessary control information) in the first transfer. Contention allows GF transmissions to be pre-allocated over dedicated resources or shared among several users. The former is more suited for fixed-pattern traffic, whereas the latter is more resource-efficient and flexible, particularly in the case of random traffic. GF transmissions over shared resources are at risk of colliding with other users who are transmitting at the same time, affecting transmission reliability. Academic research and standards groups are currently debating techniques to increase the supported load with GF random access while maintaining high reliability and low latency. GF transmissions with K-repetition, in which a pre-defined number of replicas are transmitted, and proactive repetition, in which the transmission is proactively resent until an acknowledgment is received, are examples of state-of-the-art solutions [18].

5.1 Sensing Protocol

Any UE sends an orthogonal sequence as part of their PUSCHs, and if it detects any other sequence from other UEs, it waits for a time gap before deciding what to do next. As the acpUE compete for the same resources, this process entails actual conflict. This is also a full duplex approach because the UE is transmitting its sequence while also listening to other UEs. This could cause interference, but because the UE sequences are orthogonal, there is no such disruption between the sent and received sequences.

5.2 Implementation

We take into account sensing-based collision avoidance strategies in the suggested approaches. A UE transmits an orthogonal sequence using these techniques as a component of the PUSCH Transport Block (TB). The UE can stop broadcasting the data if it discovers any other sequences being transmitted by other UEs while it is transmitting its own sequence. The UE need a time window to carry out this function in order to detect, analyze, and choose whether to delay or start the data transmission. The sequence can be based on an existing DMRS system, and it can be used to estimate the channel for data transfer. The three common mode of implementation or transmission are simplex, half-duplex,

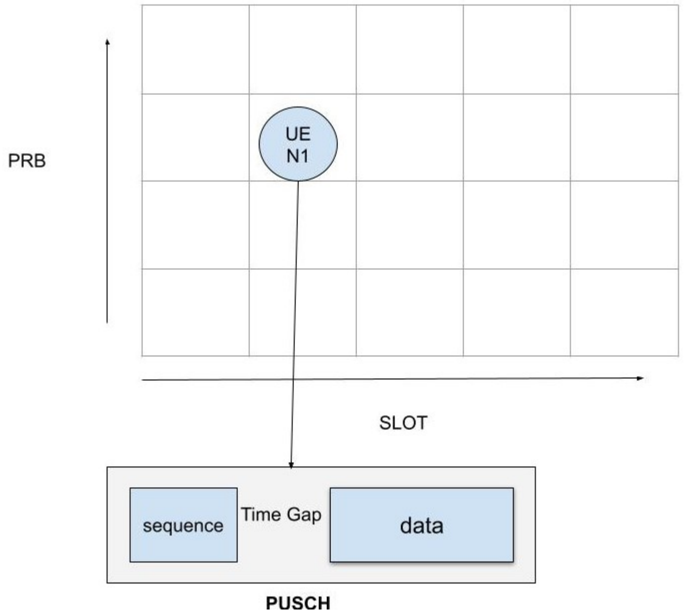


Figure 5.2. *PUSCH structure showing the three parts, sequence, time gap and data.*

and full-duplex. The direction of data transmission between any connected gadget or device is determined by the transmission mode. The difference between the two modes is that the former supports on one way communication whereas the later supports communication or transmission in both the sides but only one can perform transmission at a time. In the full-duplex a common channel is shared between the connected devices which can communicate between each other.

5.2.1 Half-Duplex

In half-duplex transmission, the two devices can transmit data in both the directions but only one at a time.

Half-duplex is similar to a walkie talkie where a person can speak at a time and the other has to listen while the other person can only speak once the person talking stops.

5.2.2 Full-Duplex

In full-duplex mode both the parties can transmit and receive at the same time through a same channel. There is no such limitation as in half duplex where one

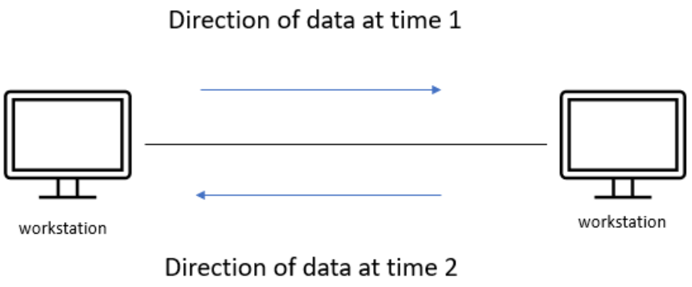


Figure 5.3. *Representation of half-duplex mode.*

has to wait for other to stop transmitting.

It is similar to being talking on a phone where there are no restrictions and there is a continuous transmission and reception of signals between the two users.

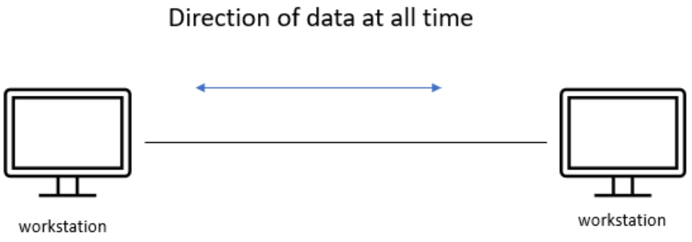


Figure 5.4. *Representation of full-duplex mode.*

The two proposals with sensing mechanism discussed in this paper are,

1. Grant-free reattempt
2. Grant-based retransmission

5.3 Grant-Free Reattempt

When another sequence in the same resource is detected, UE switches to the next slot on a random PRB and sends the identical packet without sensing. The retransmission of packets gives the user another opportunity to select a PRB where no other UE is transmitting. If two UEs try to transmit in the same PRB even in this slot, a collision can occur, and there are no alternative options for retransmission. This is a quick and efficient method for reducing collisions.

Figure 5.5. *Possible cases with two UE in a conditional grant-free reattempt.*

As illustrated in Fig. 5.5, the two UE $N1$ and $N2$ choose a same PRB in a slot 2, in this case the two UE sense the sequence of other UE and back-off to avoid the collision. The UE retransmit their data in the consecutive slot 3, but without sensing. If the PRB chosen by them is not same the data is successfully transmitted otherwise the data collides and there is a failure in transmission.

The probability of detecting other UEs' sequences in the first attempt can be approximated by Eq. (5.5). If there is detection in the first attempt, then UE does the second attempt without sensing in the next slot. So, in the second attempt all the collided UE participate again. This is an approximated model with an assumption that W is small. Hence, the load will be $\mathbb{P}_{\mathrm{c}}\lambda$ and the packet failure probability is $\mathbb{P}_{\mathrm{f}'}(W_{W=1}, \mathbb{P}_{\mathrm{c}}\lambda)$.

The probability of successful transmission from Eq. (4.5) is

$$\mathbb{P}_{\mathrm{s}}^{\mathrm{GF}} = \mathbb{P}_{\mathrm{s}} + \mathbb{P}_{\mathrm{c}}(1 - \mathbb{P}_{\mathrm{f}'}).$$

Then the failure probability of the system is

$$\mathbb{P}_{\mathrm{f}}^{\mathrm{GF}} = 1 - \mathbb{P}_{\mathrm{s}}^{\mathrm{GF}}.$$

5.4 Grant-Based Retransmission

When another sequence in the same resource is detected, UE backs off in the current slot and the base station provides a dedicated grant to UE, ensuring that no additional collisions occur. It also raises the system's average latency because the grant requires an extra slot in comparison to the conditional GF case. This sensing mechanism ensures that no collisions occur, although at the cost of additional resource consumption.

Figure 5.6. *Different scenarios that can occur in a conditional grant-based retransmission when there are two UE.*

When two UE $N1$ and $N2$ chose the identical PRB in slot 2, as shown in Fig. 5.6, there is a chance that both UE recognize each other's sequence and back-off, waiting for a grant. There will be no collision if either of the UEs can feel the other UE. Collision can happen only if both UEs are unable to perceive each other and transmit their data at the same time, however this is a highly unlikely case, making this technology reliable. The probability that a grant is successful is $\mathbb{P}_{\mathrm{G}} = 0.999$ and from Eq. (4.4) the probability that there is no error in uplink channel after a grant is successful is $1 - \mathbb{P}_{\mathrm{BER}}$.

Then the probability of successful transmission using Eq. (5.5) and (4.5) is

$$\mathbb{P}_{\mathrm{s}}^{\mathrm{GB}} = \mathbb{P}_{\mathrm{s}} + \mathbb{P}_{\mathrm{c}}\mathbb{P}_{\mathrm{G}}(1 - \mathbb{P}_{\mathrm{BER}}).$$

Then the failure probability of the system is

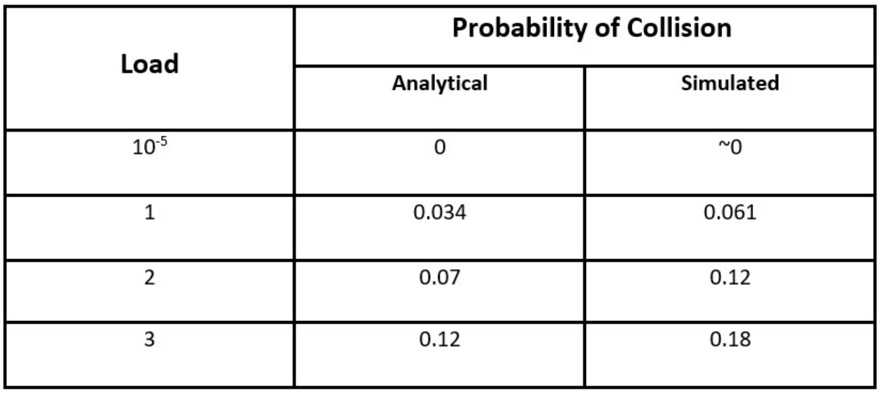
$$\mathbb{P}_{f}^{\mathrm{GB}} = 1 - \mathbb{P}_{\mathrm{s}}^{\mathrm{GB}}.$$

5.5 Probability of collision

The probability that the UEs do not collide is $\mathbb{P}_{-\mathrm{c}} = \sum_{n=1}^{\infty} \mathbb{P}_{\mathrm{n}} \left(\frac{WR-1}{WR}\right)^{n-1}$.

And the collision probability as derived in earlier section is $\mathbb{P}_{\mathrm{c}} = 1 - \mathbb{P}_{0} - \mathbb{P}_{-\mathrm{c}}$.

The table below gives the values of collision probability for no repetition schemes through analytical and simulation methods. Also, in later part the curve drawn shows that the two results follow the same trend. The analytical model shows a relative reduced collision rate at larger loads, which is inconsistent and can be studied in future works.

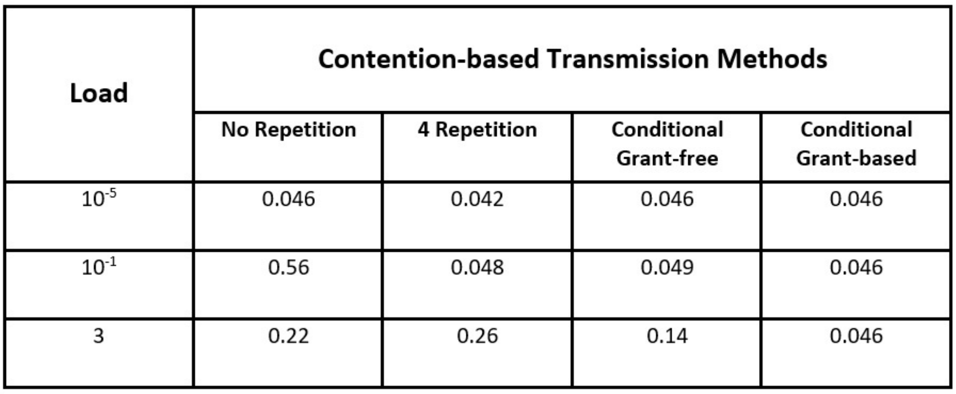


Load	Probability of Collision	
	Analytical	Simulated
10^{-5}	0	~0
1	0.034	0.061
2	0.07	0.12
3	0.12	0.18

Figure 5.7. *Comparison between analytical and simulated results.*

5.6 Probability of error

The table shows the comparison between failure probability of the four methods discussed. As more repeats aid to reduce channel errors, it can be seen that for lower load, multiple repetition method shows good performance as compared to other schemes. For low load collision error is negligible. In compared to the four repetitions, which has the added benefit of having an extra two to three repetitions, the other three schemes showed outcomes that were equivalent but performed a little worse.



Load	**Contention-based Transmission Methods**			
	No Repetition	**4 Repetition**	**Conditional Grant-free**	**Conditional Grant-based**
10^{-5}	0.046	0.042	0.046	0.046
10^{-1}	0.56	0.048	0.049	0.046
3	0.22	0.26	0.14	0.046

Figure 5.8. *Comparison of Probability of error.*

In contrast to low load performance, BER for four repeats scheme performance degrades as mean load rises, with the worst performance occurring at mean load 3. The performance of repeated repetitions degrades because transmission diversity benefits are outweighed by collision-related errors. At higher load, such as beyond mean load 0.1, the sensing-based systems outperform no-repetition and multiple repetition strategies. Only impacted UEs whose transmissions have failed in their initial attempts will retransmit in sensing-based processes. Unlike the multiple repeats approach, all UEs are not needed to transmit the same data again. As a result, there are fewer collisions in the contention pool due to less traffic congestion.

6. SIMULATION RESULTS

The Monte Carlo simulation with 10,000,000 iterations is used for the performance analysis. We used a five slot pool with three PRBs in each slot. The system uses Rayleigh fading channels, and the channel coefficients are generated using a complex normal with a mean and variance of one. The Signal-to-Noise Ratio (SNR) is set to ten decibels (dB). The channels are filled with 0.5 power additive white Gaussian noise (AWGN). Fig. 6.1 shows a comparison of the BER of the four techniques. It can be shown that the four repeats provided higher BER performance with lower load because the repetitions aid to reduce channel faults.

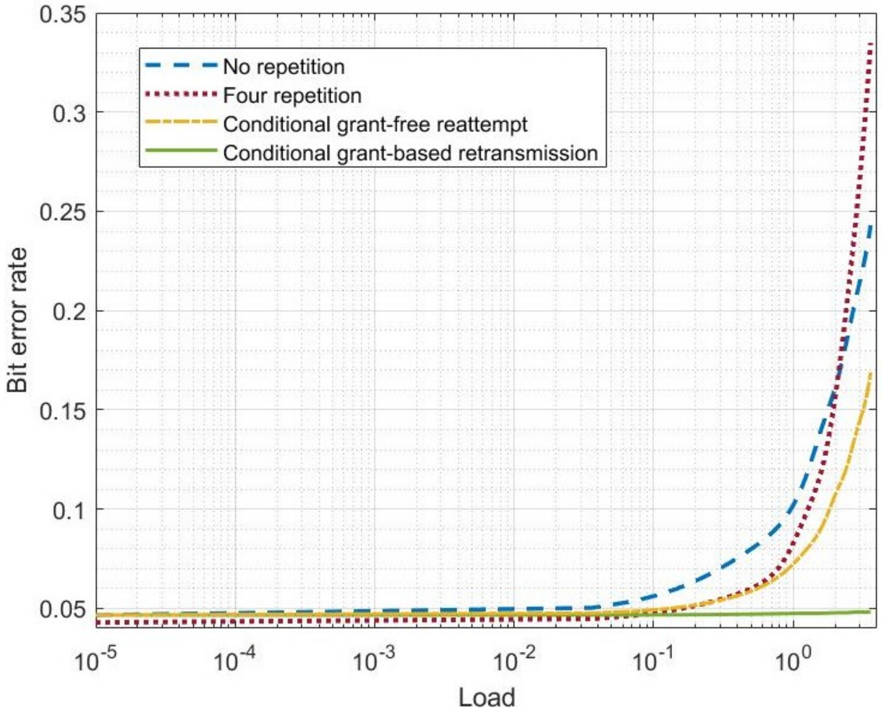


Figure 6.1. *BER comparison for the proposed and reference schemes for Poisson mean load with contention resource pool of size W = 5 slots and R = 3 PRBs per each slot.*

The other three approaches produced similar results since collisions have no ef-

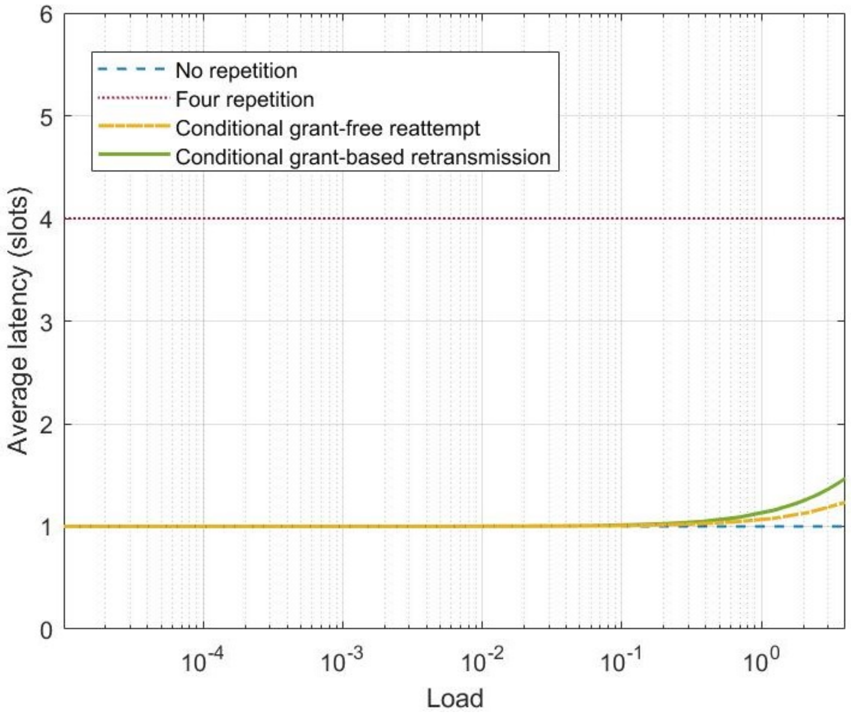


Figure 6.2. *Comparison of average latency for the four methods when simulated in a similar environment.*

fect and the error is generated solely by channel limitations. With a higher load, the BER for four repetitions rises as the volume of traffic rises, resulting in more collisions. When compared to simple systems, sensing strategies produce good outcomes for higher loads. The performance of four repetition is equivalent to conditional grant-free reattempt for mean loads between 10^1 and 10^0, but the latter is better because it requires less transmissions and processing power.The conditional grant-based strategy outperforms the other three when it comes to heavier loads.

The average delay for the four approaches can be seen and compared in Fig. 6.2. The no repetition scheme always uses a single slot but has a high BER, whereas the four repetitions scheme has a maximum latency of four, which may not be acceptable in compared to other low latency systems. When compared to conditional grant-based retransmission, the conditional grant-free technique offers a better latency performance.

The average resource utilization of the conditional grant-based technique is higher than the conditional grant-free method in Fig. 6.3, because it requires an extra

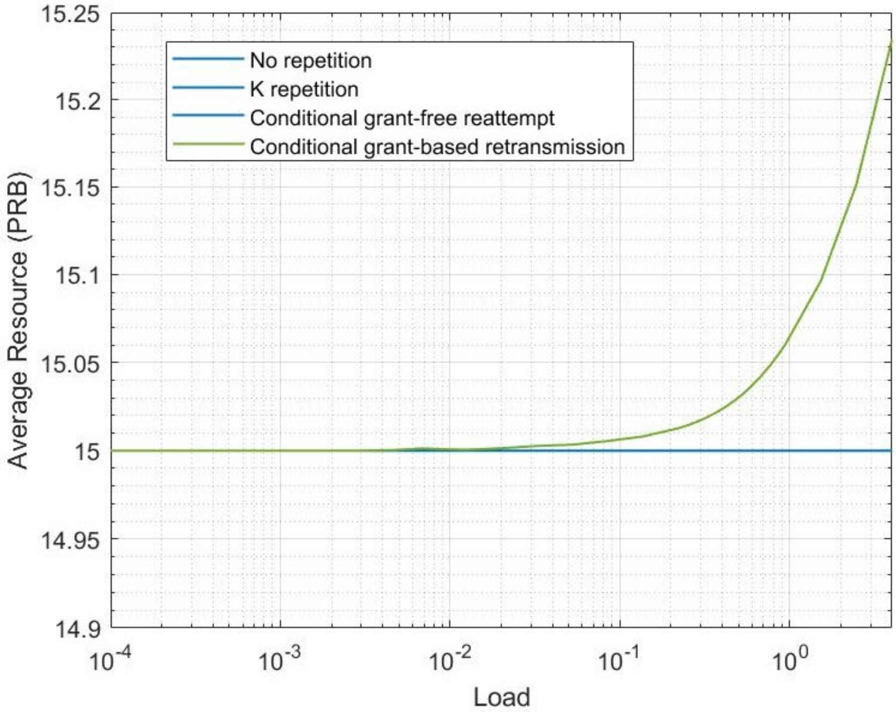


Figure 6.3. *Comparison of average resource usage for the four methods when simulated in a similar environment.*

resource each time a collision is detected in the form of the control channel used by the base station for delivering grant.

In Fig. 6.4, the maximum load is three for mean loads less than 10^1; in such instances, contention may not provide much benefit because the number of PRBs in a slot is also three. As a result, PRBs can be preassigned to UEs, and CG may be a superior option in such cases. Sensing-based contention access increases BER when the mean load is larger than 10^1, as only chosen UEs require reattempts or retransmissions due to identified collisions. One slot is utilized by the no repeats approach. It experiences the least radio delay. In the four repetitions scheme, the user equipment sends four repeats in a row. There may be one to four slots of lag. The maximal limit with four slots is shown in the figure. As they do not experience collisions, the sensing-based techniques have a delay similar to one slot at lower loads. The pool's collisions become more noticeable as the mean load rises, and the average radio latency also begins to rise. Low latency data transfers with a defined collision probability are an advantage of the contention pool. for a specific load and packet arrival rate. The various scheduling techniques, such as dynamic grant or CG, may, in some circumstances, offer substantially better latency, re-

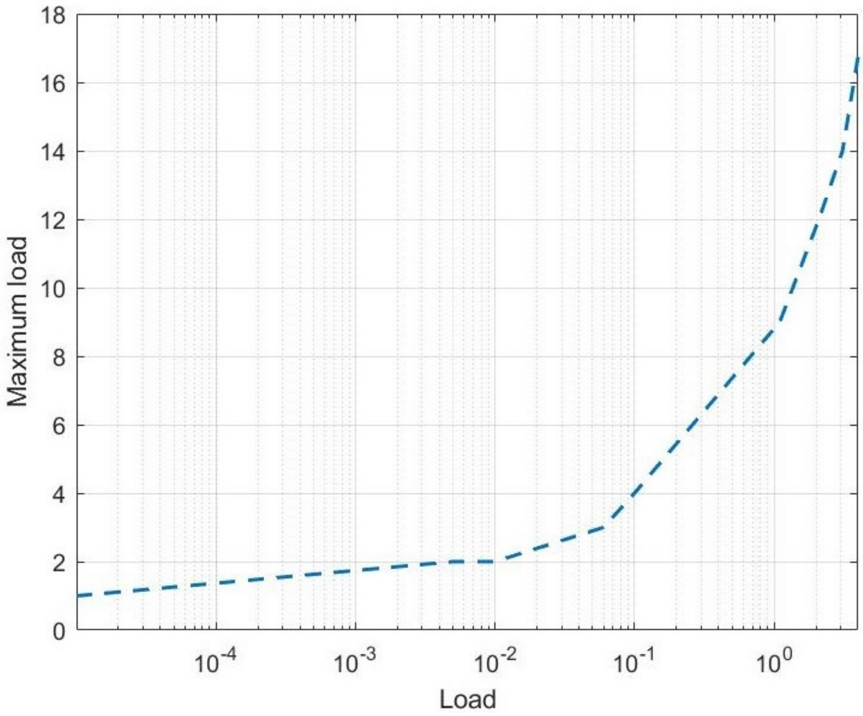


Figure 6.4. *Maximum load is depicted for Poisson mean load during the Monte Carlo simulation.*

source usage, or reduced collision rates for the same inputs. The contention pool performs worse than dedicated CG allocations when the mean load is less than 10^1. Because there are just three UEs in such cases, it is possible to allot the UEs with specific unit periodicity CGs.

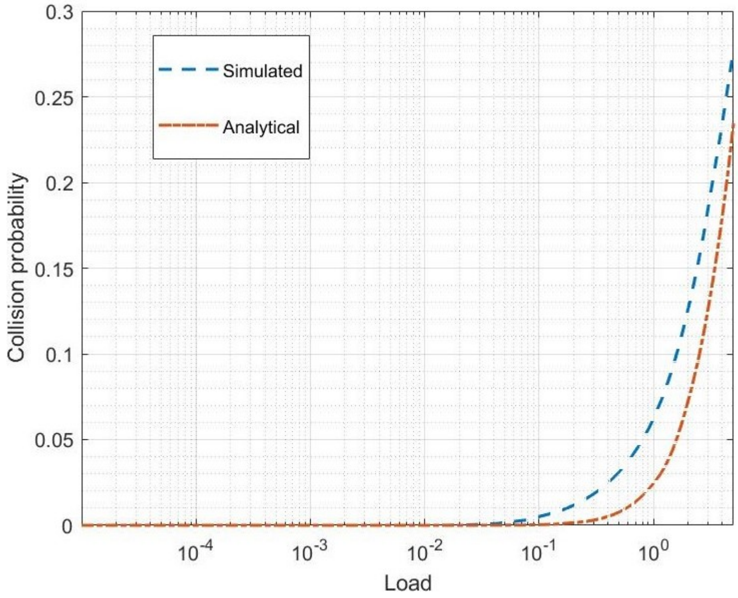


Figure 6.5. *Comparison of analytical and simulated collision probability for no repetitions scheme for Poisson mean loads*

Finally, Fig. 6.5 compares simulation-based and analytically calculated collision probability for a no repeats scheme. The same trend is followed by both curves. The analytical model is inconsistent since it shows a relative lower collision rate at larger loads, which may be further studied in other publications.

7. CONCLUSION AND FUTURE WORK

URLLC's main design goal is to provide resource-efficient wireless communication with minimal latency and excellent reliability. GF transmissions, in which users commence a transmission without the network's authorization, have emerged as a viable latency-reduction solution. However, because of its disorganized character, it is exposed to potential collisions with other users, endangering its reliability. This needs the development of improved GF schemes that can provide great dependability without losing efficiency. As discussed in several articles and papers the contention based pool has shown better results for sporadic or random traffic.

Two unique scheduling systems involving sensing were described in this work [19]. For higher mean load for low latency applications, conditional grant-free reattempt and conditional grant-based retransmission performed better than no repetition and multiple repetition in terms of bit error rate. The first approach calls for a sensing-based transmission reattempt within the same contention pool, whereas the second calls for a dedicated grant outside the pool. Although the conditional grant-free reattempt has yielded certain benefits, there is a chance that the users will collide.The second solution, conditional grant-based retransmission, totally avoids collisions at the expense of increased resource utilization and latency because of control channel transmission. We've also discovered that if the maximum load is smaller than the number of physical resource blocks, we can assign the physical resource blocks to users, resulting in zero percent collision.

The schemes could help support the shared resource structure in 6th generation technologies for networked operation mode. The schemes have an implementation cost since users must be supplied with sensing capability, and they also necessitate adjustments to the uplink shared channel design and transmission mode implementation, which can be half-duplex or full-duplex. The sensing behavior isn’t a new feature; it’s already available in the 5th generation New Radio Unlicensed for customers operating in the 5-6 GHz license-free spectrum range.

REFERENCES

[1] 3GPP, “Study on scenarios and requirements for next generation access technologies,” TS 38.913 V15.2.0, Tech. Rep., Jun. 2018.

[2] 3GPP, “Physical channels and modulation,” TS 38.211 V15.10.0, Tech. Rep., Dec. 2021.

[3] C. D. Alwis *et al.*, “Survey on 6G frontiers: Trends applications requirements technologies and future research," *IEEE OJ-COMS*, vol. 2, pp. 836-886, 2021.

[4] 3GPP, “Physical layer procedures for data,” TS 38.214 V15.9.0, Tech. Rep., Mar. 2020.

[5] G. Berardinelli *et al.*, “Reliability analysis of uplink grant-free transmission over shared resources,” in *Proc. IEEE VTC Spring*, vol. 6, pp. 23602-23611, Apr. 2018.

[6] N. H. Mahmood *et al.*, “Uplink grant-free access solutions for URLLC services in 5G new radio,” in *Proc. ISWCS*, pp. 607-612, Aug. 2019.

[7] B. Singh *et al.*, “Contention-based access for ultra-reliable low latency uplink transmissions,” *IEEE Wireless Commun. Lett.*, vol. 7, no. 2, pp. 182-185, Apr. 2018.

[8] Y. Liu *et al.*, “Analyzing grant-free access for URLLC service,” *IEEE J-SAC*, vol. 39, no. 3, pp. 741-755, 2021.

[9] T. N. Weerasinghe *et al.*, “Priority enabled grant-free access with dynamic slot allocation for heterogeneous mMTC traffic in 5G NR networks,” *IEEE Trans. Commun.*, vol. 69, no. 5, pp. 3192-3206, May 2021.

[10] C. Wang *et al.*, “Performance evaluation of grant-free transmission for uplink URLLC services,” in *Proc. IEEE VTC Spring*, pp. 1-6, Jun. 2017.

[11] S. Ozaku *et al.*, “Adaptive repetition control using terminal mobility for uplink grant-free URLLC," in *Proc. IEEE VTC Spring*, pp. 1-5, May 2020.

[12] M. C. Lucas-Estan and J. Gozalvez, “Sensing-based grant-free scheduling for ultra reliable low latency and deterministic beyond 5G networks," *IEEE Trans. Veh. Technol.*, Jan. 2022.

[13] B. Singh, “Resource Allocation for 5G Systems," PhD dissertation, School Elect. Eng., Aalto Univ., Espoo, 2021.

[14] E. Dahlman, S. Parkvall and J. Skold, ’"4G, LTE-Advanced Pro and The Road to 5G,", 2016.

[15]]T. Chapman, E. Larsson, P. Von Wrycza, E. Dahlman, S. Parkvall and J. Skold, ’"HSPA Evolution: The Fundamentals for Mobile Broadband,", 2014

[16] Z. Li *et al.*, "On contention-based 2-step random access procedure," in *Proc. WCSP*, Oct. 2020.

[17] C. Liang, S. Xia, X. Han and P. Hao, "Configured grant based URLLC enhancement for uplink transmissions," *Proc.IWCMC*, 2020.

[18] N. H. Mahmood, *et al.*, "Uplink grant-free access solutions for URLLC services in 5G new radio," in *Proc.ISWCS*, 2019.

[19] S. Seth and B. Singh, “Sensing based contention access for 6G low latency networks,”*Proc.EUCNC*, Jun. 2022.