

An Efficient Resource Allocation Algorithm for Traffic of Content Streaming in Non-Standalone OFDM Based 5G NR

Narantuya Vandantseren*, Chuluunbandi Naimannaran**, Tuyatsetseg Badarch**, Erdenebayar Lhamjav**, Otgonbayar Bataa**

*Mongolian National Defense University

**School of Information and Telecommunication Technology of MUST, Ulaanbaatar, Mongolia

vnarantuya0123@gmail.com, chuluunbandi@must.edu.mn, tuyatsetseg.b@must.edu.mn, otgonbayar_b@must.edu.mn, erdenebayar.l@must.edu.mn

Abstract— The 5G NR network provides support for a wide range of service types, including content traffic through large number of wireless connections. To ensure efficient resource allocation for specific service types, such as video streaming, a Traffic Differentiator stage is implemented at the user level. This stage segregates content queues from active users into distinct service queues, enabling tailored resource allocation. In order to investigate the performance of differentiated resource allocation in a content delivery scenario, this research paper focuses on the physical and cross-layer resource allocation of the non-standalone 5G NR network. The paper introduces a Pseudo-Inverse-based Traffic Differentiator algorithm in the TD Scheduler stage, which aims to allocate optimal radio resources to real-time and non-real-time services, while allocating the remaining resources to background services. By segregating users based on their requested services and prioritizing them differently within the service priority-specific queues, the Traffic Differentiator algorithm enhances throughput fairness among all users using Pseudo-Inverse learning. Additionally, the FD Scheduler stage utilizes our proposed optimal Channel Quality Indicator (CQI) selection algorithm to leverage Frequency Domain (FD) Multi-User (MU) diversity for resource allocation.

Keywords— Non-standalone 5G NR, Resource Allocation, Physical and Cross layer, Traffic, Content Delivering.

I. INTRODUCTION

The Third Generation Partnership Project (3GPP) has set the initial specifications for 5G New Radio (NR), which are currently considered the most widely recognized standard for 5G cellular technology. As per 3GPP, the term "5G" encompasses any system that incorporates 5G NR software, a definition that gained significant acceptance in late 2018.

The Non-Standalone (NSA) standards play a significant role in the widespread use of 5G NR. In this approach, the control plane is managed by the existing 4G LTE network, while the user plane incorporates 5G NR. The implementation of this particular standard by 3GPP aligns with the industry's goal of expediting the deployment of faster 5G services, while making use of the already established infrastructure of the existing 4G LTE network.

The latest generation of mobile networks, known as 5G New Radio (NR), has been standardized by building upon the existing concepts and features of 4G/4G+. The objective of this advancement is to bring in a greater level of adaptability, expandability, and effectiveness to the network.

Time division duplex (TDD) plays a crucial role in 5G NR by effectively managing the imbalance between uplink (UL) and downlink (DL) traffic. It achieves this by utilizing real-time traffic estimation and providing greater flexibility in allocating network resources. Just like LTE, the physical layer of 5G NR consists of essential physical channels and signals that are vital for the smooth functioning of the network [1]. This technology specifically caters to use cases such as enhanced Mobile Broadband (eMBB) and Ultra-Reliable and Low Latency Communication (URLLC), emphasizing its commitment to delivering superior broadband experiences and ensuring highly reliable, low-latency communication services.

Unlike LTE [12], NR dual connectivity involves nodes from two different Radio Access Technologies (RATs): the gNB and the eNB. The gNB provides NR access while the eNB provides E-UTRA/NR access. This creates a strong interworking between the two radio technologies and allows for a gradual integration of NR into existing LTE networks. The NR access network can operate in two modes: non-standalone and standalone. Non-standalone operation enables NR to connect within existing LTE networks, which helps expedite the deployment of 5G. On the other hand, standalone operation requires NR to connect to the

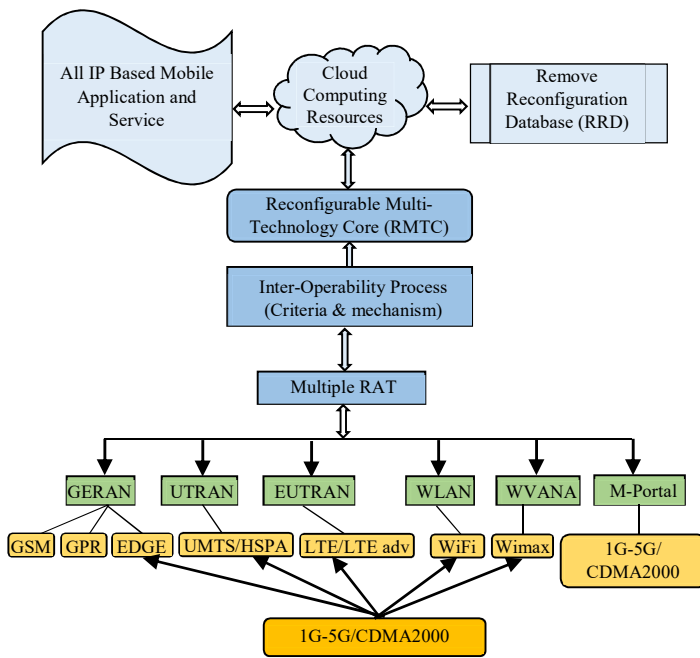


Figure 1. 5G NR system connectivity architecture

5G Core (5GC) in addition to LTE's connection to the 5GC. More detailed information about these modes can be found in the subsequent subsection [2]. The focus of this article is on the Efficient Resource Allocation Algorithm created for Non-standalone 5G NR, with a specific emphasis on the physical downlink and uplink control channels and signals.

The system connectivity architecture is depicted in Figure 1, showcasing the extensive connectivity of 5G NR, which aims to enable communication not only between humans but also between machines. It is expected that current technologies like OFDMA will continue to be effective for the next 50 years, and there is no pressing need for modifications, as stated in [3, 4].

Figure 2 depicts the Service-based architecture of the 5G NR non-standalone system, while Table 1 describes its corresponding functions. As the 4G technology features improved as an advance version of 3G in terms of speed, data bandwidth & improvement of used technology. As research work in mobile communication is focusing on 5G technology and researches are progressing towards the World Wide Wireless Web (WWW), Dynamic Adhoc Wireless Networks (DAWN) & totally real wireless world. We are expecting that 5G will be introduced in communication by 2020 which is basically user oriented. In that user can avail specific feature of 5G such as very high speed & massive data bandwidth at the low cost per bit [3]. With the help of these advanced technologies, LTE networks can now cater to a diverse range of applications and fulfill the data rate requirements set by 3GPP, which is 100 Mbps for DL and 50 Mbps for UL, along with 1.5bps/Hz spectrum efficiency in a 20 MHz bandwidth [4].

TABLE 1. 5G SYSTEM SERVICE-BASED ARCHITECTURE WITH CORE NETWORK FUNCTIONS.

		Main functions
NSSF	Network Slice Selection Function	Selects the Network Slice Instance (NSI) based on information provided during UE attach.
NEF	Network Exposure Function	Facilitates, robust, developer-friendly access the exposed network services.
NRF	Network Repository Function	Facilitates, robust, developer-friendly access the exposed network services.
UDM	Unified Data Management	Authentication Credential Repository, Access Authorization
AUSF	Authentication Server Function	Authentication and Authorization
PCF	Policy Control Function	Ensures policy charging control, authorized QoS.
AMF	Access and Mobility Management Function	NAS Signaling Termination Mobility Management Network Slicing.
SMF	Session Management Function	Selection and control of UP function, UE IP address allocation and management.
UPF	User Plane Function	Packet routing and forwarding QoS handling.
SMF	Session Management Function	Responsible for interacting with the decoupled data plane, creating updating and removing PDU sessions and managing session context with the UPF

The end-to-end network architecture of 5G, as shown in Figure 2, consists of various components. These include the Next Generation RAN (NG-RAN), also known as Cloud RAN, which is responsible for providing wireless connectivity. Additionally, there is Multi-Access Edge Computing (MEC) within the Virtual Evolved Packet Core (vEPC), which enables efficient processing at the network edge.

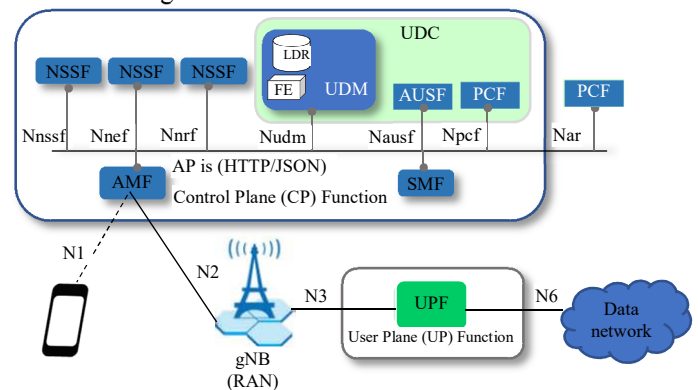


Figure 2. 5G system service-based architecture

To implement the architecture of 5G NR, foundational elements such as NS, NFV, NFV Management and Orchestration (MANO), and SDN are utilized. SDN allows for a logically centralized control and enables the programming of network devices.

The arrangement of the paper is as follows: Section II presents the physical and cross layer resource allocation design of Non standalone 5G NR system. Section III reviews the research related analysis based on the literature of the resource allocation design for the system. Section IV focuses on the proposed Pseudo Inverse scheme based Traffic Differentiator algorithm, its time and frequency scheduler sub-algorithms for content delivering of 5G NR service-oriented deployment. Finally, the research is concluded and future work is discussed.

II. PHYSICAL AND CROSS LAYER RESOURCE ALLOCATION DESIGN FOR NON STANDALONE 5G

A. 5G NR frame structure

Resource Allocation plays a crucial role in 5G NR wireless networks as it serves as the primary foundation. The efficient and equitable implementation of RA schemes heavily relies on the energy consumption and allocation of spectrum. Given the utilization of a new spectrum, the emphasis on spectrum sharing becomes paramount in the context of RA within 5G networks.

The 5G NR technology is capable of supporting both frequency division duplexing (FDD) and time division duplexing (TDD). A 5G NR frame structure is designed for efficient support of users with highly diverse service requirements. While the 5G NR frame structure shares similarities with LTE, it brings about significant improvements and important modifications. Just like LTE, each frame lasts for 10 ms and consists of 10 subframes per frame, with each subframe having a duration of 1 ms. Within a slot, there are 14 OFDM symbols, as shown in Figure 3, which highlights the structure of the 5G NR frame.

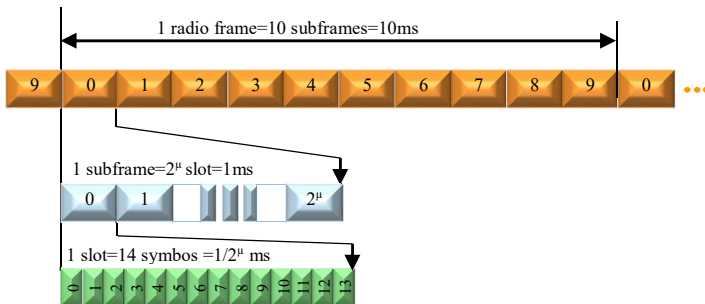


Figure 3. 5G NR Frame structure

In contrast to LTE, 5G NR introduces a significant change in the number of slots per subframe (referred to as 2^μ), which determines the duration of each slot and depends on the subcarrier spacing (SCS). Unlike LTE, where the number of slots per

subframe was always two, 5G NR allows for variability in this number. More specifically, the number of slots per frame is determined by the subcarrier spacing and can range from 15 to 240 kHz. Table 2 provides a comprehensive overview of the various choices available, along with the corresponding carrier frequencies they are specifically tailored for. Just like LTE, a Resource Element (RE) in this context refers to a single subcarrier within an OFDM symbol. However, unlike LTE, a Resource Block (RB) consists of 12 subcarriers within a single OFDM symbol (Figure 3). Moving on to Table 2, it presents the number of RBs based on the system bandwidth and subcarrier spacing, specifically for frequencies below 6 GHz.

The 5G NR technology employs a versatile subcarrier spacing (SCS) that is derived from the standard 15 KHz used in LTE. This SCS can be adjusted to values of 30, 60, and 120 KHz. In the case of a 15 KHz SCS, each subframe consists of a single slot with a duration of 1 ms. However, for a 30 KHz SCS, each subframe is divided into two slots, each lasting 500 μ s [6-8].

TABLE 2. NUMBER OF SLOTS PER SUBFRAME, SLOT DURATION, NUMBER OF SLOTS IN A FRAME AND GUARD PERIOD FOR REFERENCE SCS.

SCS	μ	Number of slots per subframe	Slots duration	Number of slots in a frame	Guard Period
15 KHz	0	1	1 ms	10	Normal
20 KHz	1	2	500 μ s	20	Normal
30 KHz	2	4	250 μ s	40	Normal/Extended
60 KHz	3	8	120 μ s	80	Normal

Each slot consists of either 14 OFDM symbols or 12 OFDM symbols, depending on whether it follows the normal Guard Period (GP) or the extended GP. Nevertheless, mini slots (2, 4, or 7 symbols) can be assigned for shorter transmissions. Additionally, slots can be combined for longer transmissions.

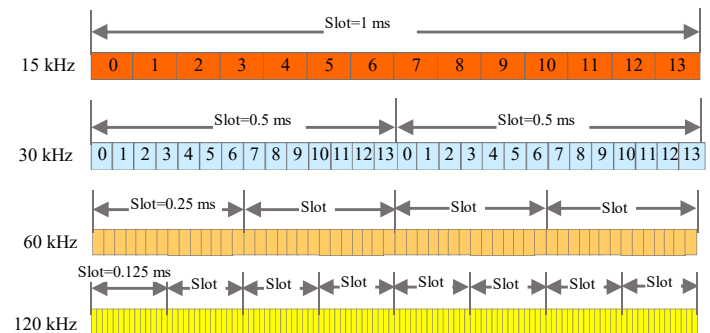


Figure 4. 5G NR scalable slot duration

The 5G NR frame structure is designed to support both Frequency Division Duplex (FDD) for paired spectrum bands and Time Division Duplex (TDD) for unpaired spectrum bands. Unlike LTE, where a subframe is made up of two slots containing 7 OFDM symbols, in 5G NR, a subframe is created by combining slots. Each slot is composed of either 7 or 14 OFDM symbols for

Sub-Carrier Spacing (SCS) of 60 kHz, and 2, 4, 7, or 14 OFDM symbols for SCS of 120 kHz. The 3GPP has standardized 255 symbol combinations, each associated with a slot format identified by a slot format index [9]. These OFDM symbols can be assigned for downlink, uplink, or flexible usage. A downlink symbol is exclusively used for downlink transmission without any simultaneous uplink transmission. Conversely, an uplink symbol is solely utilized for uplink transmission, avoiding any overlapping transmission in the downlink. Flexible symbols, however, can be adapted for transmissions in either the downlink or uplink direction.

B. 5G NR - OFDM based UL-DL pattern

3GPP has adopted CP-OFDM with a scalable numerology, incorporating subcarrier spacing and cyclic prefix, for both uplink (UL) and downlink (DL) operations, expanding this technology up to at least 52.6 GHz. Building upon LTE, 5G NR enhances the concept of carrier aggregation by introducing the Supplementary Uplink (SUL). Unlike conventional carrier aggregation, where each uplink carrier corresponds to a specific downlink carrier, SUL associates a standard downlink/uplink carrier with an additional supplementary uplink carrier operating at lower frequencies. SUL aims to extend uplink coverage and enhance uplink data rates by capitalizing on reduced path loss in low-frequency bands, particularly in scenarios with limited power availability [10].

The organization of timeslots for uplink and downlink transmission is done through DL-UL patterns. In LTE TDD, there are 7 pre-established patterns for the allocation of UL and DL in a radio frame. While there is no predefined pattern for 5G NR, a flexible pattern can be defined using parameters in TDD UL/DL Common Configuration (tdd-UL-DL-configurationCommon), as illustrated [10].

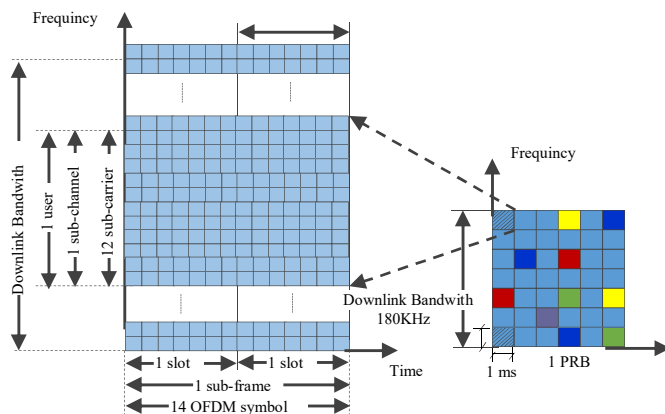


Figure 5. The OFDMA technology in 5G NR

During the air interface phase of 5G NR, mobile connectivity is made possible through the implementation of Orthogonal Frequency Division Multiple Access (OFDMA). OFDMA utilizes Orthogonal Frequency Division Multiplexing (OFDM) for each user and orthogonal subcarriers to guarantee minimal interference among users within a cell or sector. This feature is a significant advantage over previous generations such as 4G LTE [3, 11].

In the realm of 4G LTE, a Physical Resource Block (PRB) is composed of 12 subcarriers and encompasses 7 OFDM symbols within a slot duration of 0.5 ms. Conversely, in the context of 5G NR, a Resource Block (RB) is constituted by 12 subcarriers. Notably, the allocation of resources in 5G diverges from its predecessor, as it involves assigning one or more Contiguous

Resource Blocks (CRBs) to a user for a designated count of OFDM symbols.

An example of this would be the transmission of various data types, including speech, text, email, and video, to four users within a cell. To prevent any interference among them, each user is assigned specific CRBs and OFDM symbol ranges. This allocation system is dynamic and ensures that resources are utilized in an optimized manner.

5G NR utilizes two types of OFDM: cyclic prefix OFDM (CP-OFDM) and Discrete Fourier Transform Spread OFDM (DFT-S OFDM). CP-OFDM, similar to the access technology employed in LTE, incorporates a variable subcarrier spacing known as numerology, which facilitates the support of various subcarrier separations. Conversely, DFT-S OFDM, also referred to as SC-OFDM (Single Carrier OFDM), combines the transmission characteristics of a single carrier with OFDM, thereby presenting a transmission scheme akin to OFDMA. This scheme assigns distinct Fourier coefficients to different transmitters based on a complex constellation of symbols derived from the transmitted bits for each user.

The access technology employed in the second version of 5G NR is CP-OFDM (Cyclic Prefix OFDM). Unlike LTE's fixed 15kHz subcarrier spacing, CP-OFDM offers variable subcarrier spacings such as 15kHz, 30kHz, 60kHz, 120kHz, and more. This variability in subcarrier spacing results in changes in the duration of the CP symbol. The cyclic prefix, which is an integral part of OFDM schemes like OFDM, acts as a protective buffer between consecutive symbols, effectively reducing inter-symbol interference [11]. Moreover, 5G takes advantage of OFDM's combination of Quadrature Amplitude Modulation (QAM) and Frequency Division Multiplexing (FDM), enabling higher data rates for wireless networks, even in densely populated wireless network environments [13-15]. In general, a resource allocation of traffic in network area is, therefore, a challenging research area [16]. In LTE system, the algorithm of the received signal with pilot to interference cancellation algorithm for 2x2 MIMO in LTE was proposed to identify pilot-based channel estimation algorithm with low complexity [17]. The 5G NR technology

supports both Frequency Division Duplex (FDD) and Time Division Duplex (TDD) schemes, which results in the definition of NR Resource Allocation Types for both downlink and uplink specifications. These types determine how resources are allocated in the frequency domain. More specifically, the allocation methods within the downlink frequency domain are elaborated with a focus on resource allocation techniques within this domain [3].

C. Cross layer design for Traffic Differentiator

The resource allocation approach in NR closely resembles LTE scheduling, but NR demonstrates a higher level of precision, especially in its time domain scheduling within the physical layer. The suggested architecture utilizes a cross-layer concept, incorporating information from different layers to guide scheduling decisions, as illustrated in Figure 6. This approach divides the scheduling process into three stages: the Traffic Differentiator, the TD Scheduler, and the FD Scheduler, which handle the scheduling of Time Domain (TD) and Frequency Domain (FD) operations, respectively.

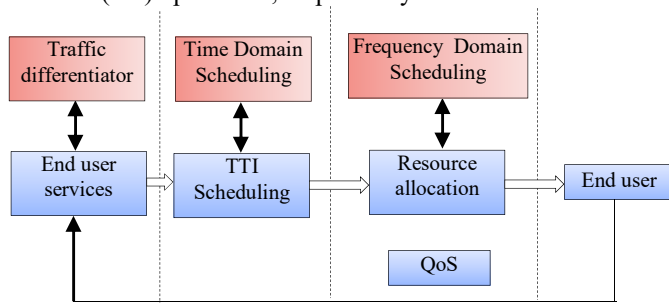


Figure 6. A cross-layer packet scheduling architecture

The Traffic Differentiator module collects information on traffic type from the application layer, Quality of Service (QoS) criteria from the network layer, queue status updates from the Radio Link Control (RLC) layer, and channel status specifics from the physical layer for each user. The TD Scheduler utilizes QoS measurements, such as average Packet Error Rate (PER) from the MAC layer, as part of the input dataset for the Pseudo-Inverse learning process. On the other hand, the FD Scheduler allocates resources across the frequency domain by taking into account user priorities and per-PRB Channel Quality Indicator (CQI) reports for efficient Physical Resource Block (PRB) mapping.

III. ANALYSIS ON RELATED WORKS

A. Analysis of Resource Allocation Algorithms for 5G NR Resource allocation

This section explores the concept of centralization in 5G NR mobile networks, with the goal of achieving adaptable processing and management that can effectively meet the demands of various services. The objective is to strike a balance between the partially decentralized networks of today and a fully centralized cloud

radio access network. To address this middle ground, the concept of Radio Access Network as a Service is introduced, which involves partially centralizing radio access network functionalities based on the specific needs of the network [18]. The discussion then may fall into various scheduling schemes that form the foundation of networks [19]. These include Round Robin, BestCQI, Proportional Fair, MLWDF, and EXP-PF, as well as newer schemes developed based on specific metrics outlined in Table 4. Round Robin, for example, is a simple method that evenly distributes resources among all users in the current cell [21], [22]. Its simplicity has made it widely used [20]. On the other hand, the Best CQI scheme prioritizes users with the best Channel Quality Indicator (CQI) [23]. Proportional Fair (PF) divides achieved data rates by past average rates, ensuring fairness by giving priority to users with lower previous rates [24], [25]. Variations of MLWDF, such as EXP-MLWDF, combine MLWDF with an exponential term to favor users with poor channel conditions [26, 27].

In addition to these schemes, there have been other developments in the field. For instance, Charles Katila's Neighbors Aware Proportional Fair (N-PF) method takes into account the presence of IoT devices adjacent to each scheduled user in heterogeneous systems [28]. Furthermore, there is a scheduler specifically designed for eMBB downlink, aiming to optimize resources and prioritize cell-edge users [29, 30]. Another proposed scheduler focuses on prioritizing URLLC flows based on the signal-to-noise ratio multiplied by a normalized queue state parameter, which helps prevent buffer congestion [31].

In the context of 3GPP LTE-Advanced, specifically in Release 10 and 11, it was acknowledged through feasibility studies that there existed sufficient capacity for further enhancements in performance. These enhancements were particularly targeted towards the cell edges and involved coordinated transmission among multiple transmitters located across different cell sites [33]. The conventional PS algorithms, such as MAX C/I and Proportional Fairness (PF) algorithms [1], solely focus on enhancing resource utilization based on the channel conditions of users. However, they do not take into account the Quality of Service (QoS) requirements, such as the delay requirements for real-time (RT) traffic or the minimum throughput requirements for non-real-time (NRT) traffic [34]. To address this limitation, several QoS-aware PS algorithms have been developed based on the Modified Largest Waited Delay First (M-LWDF) approach [35]. The M-LWDF algorithm aims to improve the packet delay and Packet Delivery Ratio (PDR) for RT traffic, as well as meet the minimum throughput requirement for NRT traffic over extended periods of time.

The M-LWDF algorithm is utilized in Frequency Division Multiplexing (FDM) systems to optimize the allocation of sub-carriers in OFDMA-based networks [37]. Within the Sum Waiting Time-Based Scheduling (SWBS) algorithm [6], M-LWDF is modified by updating the queue statuses after each

allocation of sub-carriers. SWBS prioritizes Real-Time (RT) and Non-Real-Time (NRT) traffic types based on the product of cumulative waiting times of packets and individual user channel conditions. The objective of this algorithm is to improve the Quality of Service (QoS) for RT and NRT traffic in OFDMA-based systems, showcasing enhanced QoS in comparison to M-LWDF. However, despite the improvement in QoS for RT and NRT traffic, these algorithms [34-39] do not simultaneously enhance the overall system throughput and fairness among users.

B. Analysis of Content Delivering

There are currently two primary mechanisms for sharing spectrum in use: distributed and centralized. In light of the findings obtained, a total of ten recommendations have been put forward to enhance the allocation of resources within 5G networks [40]. The non-standalone (NSA) mode in 5G NR involves deploying a system that relies on the control plane of an existing 4G LTE network for control functions, while dedicating 5G NR exclusively to the user plane [41][42]. Various algorithms for resource allocation cater to service control, such as cooperative game theory, a virtual token mechanism, and notable ones like EXP-RULE and Modified-Largest Weighted Delay First (M-LWDF) algorithms that prioritize real-time flows, particularly in the downlink system. To meet the demands of maintaining high data quality without loss or delay, several packet scheduling algorithms have emerged, including Maximum-Largest Weighted Delay First (M-LWDF), Proportional Fair (PF), and Exponential Rule (EXP-RULE) [43-45]. The best CQI scheduler assigns resource blocks (RBs) to the user with the best radio link conditions or channel quality for a particular RB at every TTI. Each cellular user sends a CQI to the eNodeB to perform the scheduling. The eNodeB transmits a reference signal (i.e., downlink pilot) in the downlink channel to the cellular user [46].

The design of 5G NR integrates LTE elements with a new radio access technology that does not support backward compatibility with LTE. It is worth noting that 5G NR functions over a broad spectrum, ranging from less than 1 GHz to 100 GHz. In the domain of 5G, PCF, which stands for Policy Control Function, operates as a Network Function (NF) that assumes the responsibility of shaping policy control and charging regulations for 5G services and applications. Through service-based interfaces, PCF interacts with various network functions such as AMF, SMF, UPF, UDM, AUSF, and AF (Table 1). Its primary role involves authorizing Quality of Service (QoS) and assigning QoS flow identifiers (QFI) for individual data flows. In addition to these functions, PCF enables network slicing, roaming, and mobility management. Moreover, it collects subscriber metrics and provides real-time management of subscribers, applications, and network resources in accordance with the business rules defined by the service provider.

The scheduling process in 5G NR, which is a part of cellular communication including LTE, is mainly controlled by the network, while the UE adheres to the instructions given. Although the scheduling mechanism in NR is similar to LTE, it has a more refined granularity, particularly in time domain scheduling at the physical layer. This increased granularity enables more accurate scheduling and management, particularly in the time domain aspects of the physical layer in NR when compared to LTE.

After conducting a thorough examination of the network architectures of 4G and 5G, a set of eight criteria has been identified for careful consideration during the resource allocation process in 5G networks [42]. These criteria, as presented in Table 3, encompass Delay, Fairness, Packet Loss Ratio, Spectral Efficiency, Throughput, Electrical Energy Usage, Processor Cycle, and Memory.

TABLE 3. CRITERIA FOR RESOURCE ALLOCATION

Criteria	Description
Delay	Delay is defined as the latency measured when packets are transmitted from a source to a destination, and it is measured in milliseconds
Fairness	Fairness is the percentage of QoS requirements met during resource allocation.
Packet Loss Ratio	Packet loss ratio is the ratio of packets received by the receiver to the number of packets sent by a sender.
Spectral Efficiency	Spectral efficiency is the rate at which information is transmitted on a provided bandwidth and it is measured in bits/sec/Hz.
Throughput	Throughput is the number of bits in a flow processed per unit of time in a network and it is measured in Kbps.
Electrical Energy Usage	Electrical energy is the electricity used by electronic devices such as routers, switches, servers, and equipment to perform their work. It is usually measured in kWh. Different devices have different needs in terms of electricity according to their functionalities within the 5G end-to-end infrastructure, so having a differentiated approach to electrical energy resource allocation is needed.
Processor Cycle	The processor cycle is the time taken to run an elementary instruction by a processor in a machine or computer. SDN and NFV have increased the need for customized high-end processing on every device along with the 5G end-to-end infrastructure so that assigning processor cores and processor cycles no longer require a one size fits all approach.
Memory	Memory is the amount of main memory or Random-Access Memory (RAM) that is available for all devices in the 5G end-to-end architecture to store running programs. Virtual machines hosting VNFs now have main memory requirements tailor-made for their functionalities.

The development of the 5G NR aims to achieve high throughput, low latency, scalability, ubiquitous connectivity, and energy-efficient solutions. The adaptability and scalability of 5G NR technology within the physical layer design provide benefits, enabling it to support various use cases, such as using orthogonal frequency division multiple access (OFDMA) at the medium access control (MAC) layer with different numerologies. The effectiveness of OFDMA depends on how the access point (AP) allocates channel resources among stations (STAs), with commonly used schedulers including round-robin (RR), maximum rate (MR), and proportional fair (PF). By improving resource allocation, the envisioned 5G NR non standalone version based services can be effectively delivered.

IV. Proposed Algorithms of content delivering for 5G NR service-oriented deployment

The mobile network's amalgamation of various service requests can be likened to a mixed-traffic flow that is directed towards the eNB. This flow showcases a range of Quality of Service (QoS) requirements. To segregate users based on their unique service requests, the Traffic Differentiator funnels them into service-specific queues that are aligned with varying scheduling priorities. Each queue employs a tailored queue sorting algorithm that caters to the specific QoS demands of the respective service type. This study presents the classification of mixed traffic into six distinct queues, with an RT queue comprising emergency message content, followed by RT and NRT queues designated for streaming video services. These queues are structured in descending priority order, with the RT queue positioned at the top, followed by the NRT queue.

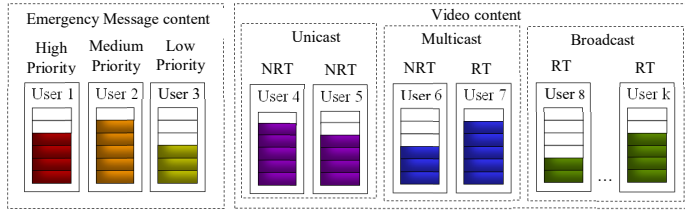


Figure 8. Traffic differentiator Stage

Emergency message content is given the highest priority status, while streaming video services are considered to be high-throughput services that can be delivered in real-time or non-real-time. The scheduling information for emergency content is transmitted before the actual data, carrying critical scheduling grants or allocated Physical Resource Blocks (PRBs) that are used for data transmission to users. In the case of non-real-time streaming video, it is crucial to maintain a sustained minimum throughput guarantee in order to ensure high-quality video streaming over extended periods of time. The transmission of emergency information to all scheduled users follows a First Come First Serve (FCFS) approach. Furthermore, two innovative Service Priority Specific queue Sorting Algorithms (SPSSA)

have been introduced—one for real-time emergency messages and another for real-time and non-real-time streaming services [47]. The prioritization of users who request background services is determined based on the priority coefficient, which aims to strike a balance between fairness and system throughput.

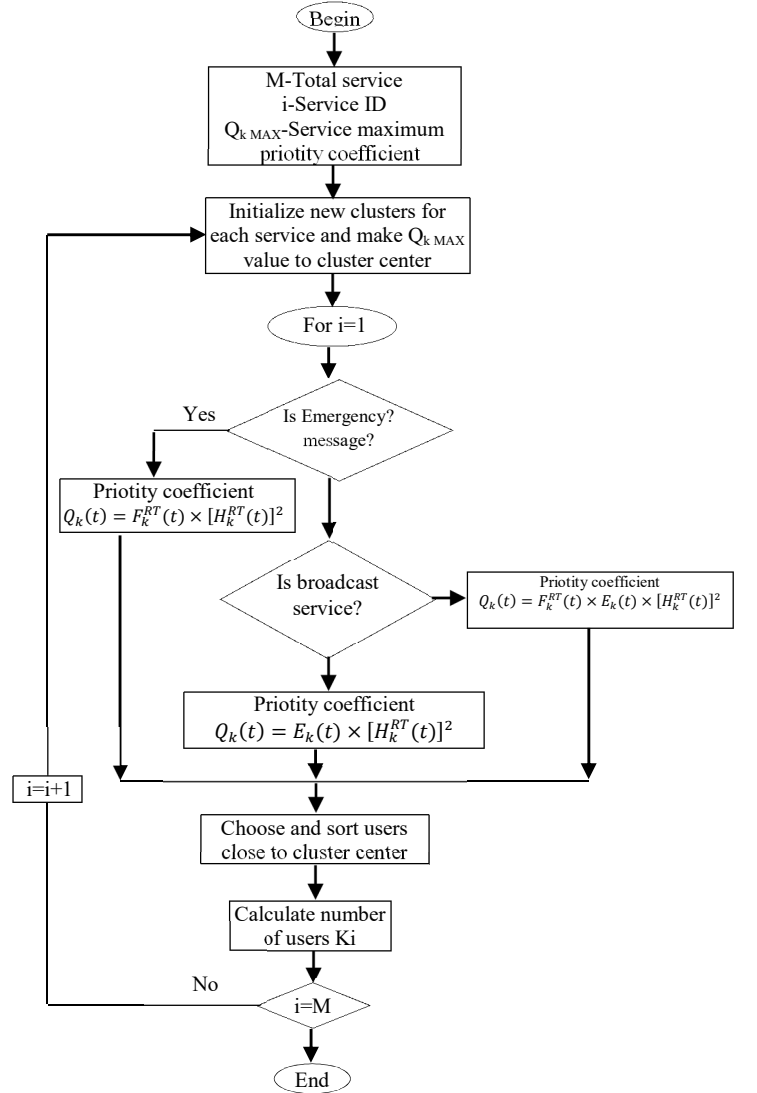


Figure 9. Traffic differentiator Stage uses Dynamic C-mean clustering algorithm

Figure 9 illustrates the complex operation of the Traffic Differentiator, demonstrating the categorization of mixed traffic into separate segments: RT emergency users, NRT streaming users, and the allocation of their control information into a dedicated queue. This procedure employs a Dynamic c-mean clustering algorithm to effectively organize the priorities of chosen RT users. The arrangement is determined by their individual $Q_k(t)$ priority coefficients, with the objective of enhancing and optimizing the priority metric to enhance system

performance. Non standalone 5G NR uses QCI to categorize different types of data traffic based on their requirements and characteristics.

A. Time Domain Scheduler

The TD Scheduler plays a crucial role in determining the users that should be scheduled in the ongoing Transmission Time Interval (TTI), taking into account both the available radio resources and the various user priorities. The TD Scheduler stage introduces a proposed Time Domain Scheduling Algorithm based on Pseudo-Inverse. This algorithm incorporates the Pseudo-Inverse process, which enables adaptive resource allocation. By applying the Pseudo-Inverse (Figure 10) structure, the algorithm identifies the optimal number of Real-Time (RT) and Non-Real-Time (NRT) users to be scheduled in the current TTI. This approach guarantees a fair distribution of resources across different service types, ultimately improving the overall system-level performance.

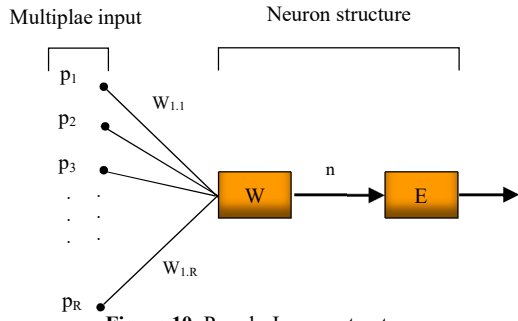


Figure 10. Pseudo-Inverse structure

In the process of pseudo-inverse, the input values are referred to as p_i , whereas the corresponding output values are designated as r . Both the input and output values are expressed as row vectors. By utilizing a methodology similar to that employed in neural networks, we determine the quantity of chosen requests from the complete set of user requests for different services during a given time period (Figure 11). This selection procedure assists in determining the ideal number of requests to be accommodated within a specific timeframe.

$$p_c = \begin{bmatrix} K_1 \\ K_2 \\ \dots \\ K_S \end{bmatrix} \quad s = 1, \dots, S \quad (1)$$

$$P = [p_1 \ p_2 \ p_c \ \dots \ p_C] \quad c=1, \dots, C \quad (2)$$

$$K_{DemandUser\ s} = \sum K \quad (3)$$

P - A total service cluster, p_c - Random service matrix, C - Total service numbers, S - Total service classify

Recall that the task of the linear associator was to produce number of users for an input of s^{th} service. The primary goal of the linear associator was to produce the tally of users linked to a given service input to furnish the quantity of users connected to a particular service input.

$$t_s = W p_s \quad (s = 1, 2, 3) \quad (4)$$

Let's rewrite Eq. (4) in matrix form:

$$T = W P \quad (5)$$

Here, t_s - s^{th} service output, W - Pseudo weight matrix, T - Selected users from each service.

A coefficient (E) determines whether the desired output value is 100% verified by the processing of multiple input values.

$$E = T - W P \quad (6)$$

The pseudo-inverse weighting matrix has defined output values depending on input multiple input values. A size of weighting matrix equal size of input matrix. Obtain Weight matrix:

$$W = T P^+ \quad (7)$$

$$P^+ = (P^T * P)^{-1} * P^T \quad (8)$$

P^+ - Moore Penrose pseudo inverse, P^T - Reverse row and column matrix, P^{-1} - Inverse matrix.

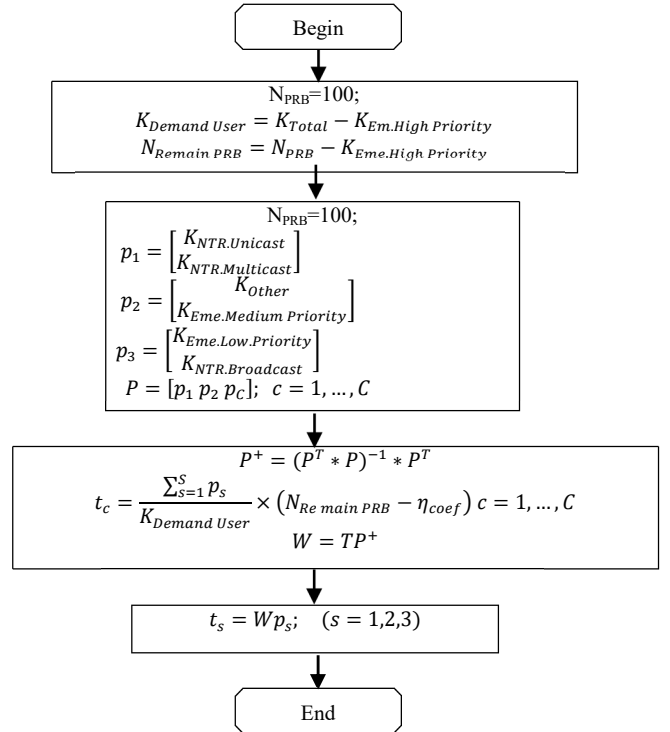


Figure 11. Pseudo-Inverse based Time Domain Scheduling algorithm

B. Frequency Domain Scheduler

The role of the FD Scheduler has a significant value as it determines the specific allocation of Physical Resource Blocks (PRBs) to the users selected by the TD Scheduler. In order to optimize the assignment of PRBs to each user, multi-user diversity is leveraged in the frequency domain. Once the resource allocation decision is made, the Quality of Service (QoS) measurement unit comes into play and calculates key metrics such as the average throughput of Non-Real-Time (NRT) users, average delay, and average Packet Error Rate (PER) of Real-Time (RT) users. This crucial information is then transmitted to the TD Scheduler stage, where it contributes to the decision-making process in the subsequent Transmission Time Interval (TTI). It is worth noting that the availability of accurate Channel Quality Indicator (CQI) reports from all users in each TTI is assumed, as commonly acknowledged in existing literature on channel-aware scheduling.

$$PER_{average} = \frac{1}{K} \sum_{k=1}^K PER_k \quad (9)$$

$$PER_k(t) = 1 - \prod_{j=1}^J (1 - BLER_j) \quad (10)$$

$PER_{average}$ - average PER, $BLER_j$ - Block error ratio for each users

The available resource blocks are allocated to user through an iterative process. At each iteration only one RB is allocated to the user which maximizes the following proposed priority function.

$$RF(k) = \frac{RSRP_i(k) \times BLER_i(k) \times P_i(k)}{RSSI_t \times PER_i \times p_i} \quad (11)$$

$RSRP_i(k)$ -k-th user power level. $BLER(k)$ – k-th user block error ratio, $P_i(k)$ - k-th user service level coefficient, $RSSI_t(k)$ -total power level of free band width, PER_i - packet error ratio

The total transmit power of an OFDM symbol in the t^{th} TTI can be expressed as:

$$RSSI_t = \sum_{r=1}^R (a_{k,r}, RSRP_i(k)), \forall r \in S_r^t \quad (12)$$

Algorithm: Optimal CQI selection algorithm	
No	Input: K, S_r, S_t, S_p , Output: $a_{k,r}, b_{k,r}$
1	Initialization
2	Set $a_{k,r} = 0, b_{k,r} = 0, t = 1, U = \{1, 2, \dots, K\}$
3	Calculate $RSSI_t$ in Eq(2)
4	Calculate $PRF(k), \forall k, \forall i \in U$
5	While $S_r^t \neq \emptyset$
6	If $U = \emptyset$ Set $S_t = S_t - t$ and GOTO Step(h)
7	for $k \in U$

8	Find $k^* = \operatorname{argmin}_k \frac{ S_r^k }{PRF(k)}$
9	Find $t^* = \operatorname{argmin}_t S_r^t \cap S_r^k , \forall \in S_t$
10	Find $r^* = \operatorname{argmax}_r S_{k^*,r^*}, \forall r \in S_r^{t^*} \cap S_r^{k^*}$
11	Set $a_{k^*,r^*} = 1, b_{k^*,r^*} = CQI_{r^*}(k^*)$
12	Set $S_r^t = S_r^t - r^*, S_r^{k^*} = S_r^{k^*} + r^*$
13	Add n to user subset $U_i = \{U_1, U_2, \dots, U_k\}$
14	if $t > S_t $ Set $t = 1$ else Set $t = t + 1$

The frequency of reporting CQI is typically set to once every few milliseconds. It is recommended in [46] that, in static channel conditions, the UE report CQI such that it achieves a block error rate (BLER) 10% when scheduled data corresponding to the median reported CQI. For example, in HSDPA, the UE monitors the quality of the downlink wireless channel and periodically reports this information to the base station (referred to here as NodeB) on the uplink [48]. The channel quality indicator (CQI) feedback mechanism provides various signaling options to enable the radio link adaptation between a user equipment (UE) and a base-station.

V. CONCLUSION

In the era of a content delivering growing traffic, the broadcast and multicast technologies will be able to enhance the efficiency of network resource utilization. Non-standalone 5G New Radio (NR) access technology in the physical layer design accommodate to be flexible and scalable that can support many use cases such as orthogonal frequency division multiple access (OFDMA) at its medium access control (MAC) layer and numerology. This research supports a fair distribution of resources across different service types, ultimately improving the overall system-level performance for non-standalone 5G NR system. The proposed algorithms of a content delivering for 5G NR service-oriented deployment focuses on the classification of mixed traffic into distinct queues, with an RT queue comprising emergency message content, followed by RT and NRT queues designated for streaming video services.

Traffic Differentiator Traffic FD Scheduler supports CQI is to determine the specific allocation of Physical Resource Blocks (PRBs) to the users selected by the TD Scheduler. This study analyzed the downlink access with varying resource allocations the number of user equipment, and numerology with additional proposed schedulers. Furthermore, the further studies focus on the ultimate results of the proposed algorithms compared to the other related algorithms. In addition, our future research will intend to get the effective result of to use AI in the TD scheduler to intelligently pick users from the queues to improve the QoS of different traffic types and overall system performance.

REFERENCE

- [1] C.-X. Wang et al., "Cellular architecture and key technologies for 5G wireless communication networks," *IEEE Comm. Mag.*, vol. 52, no. 2, pp. 122–130, Feb. 2014.
- [2] Rinaldi, F., Raschella, A. & Pizzi, S. 5G NR system design: a concise survey of key features and capabilities. *Wireless Netw* 27, 5173–5188 (2021). <https://doi.org/10.1007/s11276-021-02811>
- [3] S. Kumar, T. Agrawal, and P. Singh, "A Future Communication Technology: 5G," *International Journal of Future Generation Communication and Networking*, vol. 9, no. 1, pp. 303–310, 2016, doi: 10.14257/ijfgcn.2016.9.1.26.
- [4] Gupta A, Jha RK. A survey of 5G network: Architecture and emerging technologies. *IEEE Access*. 2015;3:1206-1232. DOI: 10.1109/ACCESS.2015.2461602
- [5] B. Sokappadu, A. Hardin, A. Mungur, S. Armoogum, "Software Defined Networks: Issues and Challenges," *Conference on Next Generation Computing Applications (NextComp)*, Mauritius, 2019.
- [6] Technical Specification: NR Physical Layer Procedures for Control. TS 38.213 V16.2.0. 2020
- [7] Kamal MA, Raza HW, Alam MM, Su'ud MM, Sajak ABAB. Resource Allocation Schemes for 5G Network: A Systematic Review. *Sensors (Basel)*. 2021 Oct 2;21(19):6588. doi: 10.3390/s21196588. PMID: 34640908; PMCID: PMC8512213.
- [8] Degambur, L. -, Mungur, A., Armoogum, S., & Pudaruth, S. (2021). Resource allocation in 4G and 5G networks: A review. *International Journal of Communication Networks and Information Security*, 13(3), 401–408.
- [9] 3GPP (2020). TS 38.213; NR; Physical layer procedures for control; Rel.16; 3GPP: Sophia Antipolis Valbonne, France.
- [10] Rinaldi, F., Raschella, A. & Pizzi, S. 5G NR system design: a concise survey of key features and capabilities. *Wireless Netw* 27, 5173–5188 (2021). <https://doi.org/10.1007/s11276-021-02811>
- [11] <https://telcomaglobal.com/>
- [12] 3GPP (2014). TR 36.842; Study on small cell enhancements for E-UTRA and E-UTRAN; Higher layer aspects; Rel.12. 3GPP: Sophia Antipolis Valbonne, France.
- [13] R. N. Mitra, P. Agrawal, "5G Mobile Technology: A Survey", *ICT Express*, Vol. 1, No. 3, pp. 132–137, 2015.
- [14] L. B. Le, V. Lau, E. Jorswieck, N-D. Dao, A. Haghighat, D. I. Kim, T. Le-ngoc, "Enabling 5G mobile wireless technologies," *EURASIP Journal on Wireless Communications and Networking*, Article 218, 2015.
- [15] B. Bangerter, S. Talwar, R. Arefi, K. Stewart, "Network and Devices for the 5G Era," *IEEE Communications Magazine*, Vol. 52, No. 2, pp. 90–96, 2014
- [16] T. Badarch and O. Bataa, "Big Data Area: A Novel Network Performance Analysis Technique Based on Bayesian Traffic Classification Algorithm," 2017 International Conference on Green Informatics (ICGI), Fuzhou, China, 2017, pp. 238–245, doi: 10.1109/ICGI.2017.46.
- [17] O. Bataa, E. Lamjav, U. Purevdorj, Y. i. Kim and K. Gonchigsumlaa, "ICS Algorithm Implementation in 3GPP/LTE," 2014 7th International Conference on Ubi-Media Computing and Workshops, Ulaanbaatar, Mongolia, 2014, pp. 17–21, doi: 10.1109/U-MEDIA.2014.41.
- [18] M. Peng, Y. Li, Z. Zhao, and C. Wang, "System architecture and key technologies for 5G heterogeneous cloud radio access networks," *IEEE Netw.*, vol. 29, no. 2, pp. 6–14, Mar./Apr. 2015. Available: <http://ieeexplore.ieee.org/document/7986957/>
- [19] T. Badarch and O. Bataa, "An adaptive scheduling scheme to efficient emergency call holding times in public safety network," *Ifost, Ulaanbaatar, Mongolia*, 2013, pp. 210–213, doi: 10.1109/IFOST.2013.6616888.
- [20] W. Mansouri, K. B. Ali, F. Zarai, and M. S. Obaidat, "Radio resource management for heterogeneous wireless networks: Schemes and simulation analysis," in *Modeling and Simulation of Computer Networks and Systems*. Amsterdam, The Netherlands: Elsevier, 2015, pp. 767–792. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/B9780128008874000274>
- [21] S. Shakkottai and R. Srikant, "Scheduling real-time traffic with deadlines over a wireless channel," in *Proc. 2nd ACM Int. Workshop Wireless Mobile Multimedia (WOWMOM)*. New York, NY, USA: ACM Press, 1999, pp. 35–42. [Online]. Available: <http://portal.acm.org/citation.cfm?doid=313256.313273>
- [22] X. Yuan and Z. Duan, "Fair Round-Robin: A low complexity packet scheduler with proportional and worst-case fairness," *IEEE Trans. Comput.*, vol. 58, no. 3, pp. 365–379, Mar. 2009. [Online]. Available: <http://ieeexplore.ieee.org/document/4626953/>
- [23] S. Schwarz, C. Mehlführer, and M. Rupp, "Low complexity approximate maximum throughput scheduling for LTE," in *Proc. Conf. Rec. 44th Asilomar Conf. Signals, Syst. Comput. (ASILOMAR)*, Nov. 2010, pp. 1563–1569. [Online]. Available: <http://ieeexplore.ieee.org/document/5757800/>
- [24] G. Femenias, F. Riera-Palou, X. Mestre, and J. J. Olmos, "Downlink scheduling and resource allocation for 5G MIMO-multicarrier: OFDM vs FBMC/OQAM," *IEEE Access*, vol. 5, pp. 13770–13786, 2017. [Online].
- [25] S. O. Aramide, B. Barakat, Y. Wang, S. Keates, and K. Arshad, "Generalized proportional fair (GPF) scheduler for LTE-A," in *Proc. 9th Comput. Sci. Electron. Eng. (CEECE)*, Sep. 2017, pp. 128–132. [Online]. Available: <http://ieeexplore.ieee.org/document/8101612/>
- [26] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, P. Whiting, and R. Vijayakumar, "Providing quality of service over a shared wireless link," *IEEE Commun. Mag.*, vol. 39, no. 2, pp. 150–154, Feb. 2001. [Online]. Available: <http://ieeexplore.ieee.org/document/900644/>
- [27] M. Mahfoudi, M. E. Bekkali, A. Najd, M. E. Ghazi, and S. Mazer, "A new downlink scheduling algorithm proposed for real time traffic in LTE system," *Int. J. Electron. Telecommun.*, vol. 61, no. 4, pp. 409–414, Dec. 2015. [Online]. Available: <http://journals.pan.pl/dlibra/publication/101933/edition/87947/content>
- [28] C. J. Katila, C. Buratti, M. D. Abrignani, and R. Verdone, "Neighbors-aware proportional fair scheduling for future wireless networks with mixed MAC protocols," *EURASIP J. Wireless Commun. Netw.*, vol. 017, no. 1, p. 93, Dec. 2017. [Online]. Available: <https://jwcn.eurasipjournals.springeropen.com/articles/10.1186/s13638-017-0875-6>
- [29] M. I. Saglam and M. Kartal, "5G enhanced mobile broadband downlink scheduler," in *Proc. 11th Int. Conf. Electr. Electron. Eng. (ELECO)*, Nov. 2019, pp. 687–692. [Online]. Available: <https://ieeexplore.ieee.org/document/8990378/>
- [30] W. S. Afifi, A. A. El-Moursy, M. Saad, S. M. Nassar, and H. M. El-Hennawy, "A novel scheduling technique for improving cell-edge performance in 4G/5G systems," *Ain Shams Eng. J.*, vol. 12, no. 1, pp. 487–495, Mar. 2021. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S2090447920301878>
- [31] P. Korrai, E. Lagunas, S. K. Sharma, S. Chatzinotas, A. Bandi, and B. Ottersten, "A RAN resource slicing mechanism for multiplexing of eMBB and URLLC services in OFDMA based 5G wireless networks," *IEEE Access*, vol. 8, pp. 45674–45688, 2020. [Online]. Available: <https://ieeexplore.ieee.org/document/9020161/>
- [32] A. Karimi, K. I. Pedersen, and P. Mogensen, "Low-complexity centralized multi-cell radio resource allocation for 5G URLLC," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, May 2020, pp. 1–6. [Online]. Available: <https://ieeexplore.ieee.org/document/9120469>
- [33] W. Nam, D. Bai, J. Lee, and I. Kang, "Advanced interference management for 5G cellular networks," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 52–60, May 2014.
- [34] B. Chisung, and C. Dong, "Fairness-Aware Adaptive Resource Allocation Scheme in Multihop OFDMA System", "Communication letters, IEEE, vol.11, pp.134- 136, Feb. 2007
- [35] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar and P. Whiting, "Providing Quality Over a Shared Wireless Link" *IEEE communications magazine*, February 2001.
- [36] S. Shakkottai, A.L. Stolyar, "Scheduling for Multiple Flows Sharing a Time-Varying Channel: The Exponential Rule" *Bell Labs, Lucent Technologies*, NJ 07974.

- [37] P. Parag, S. Bhashyam, and R. Aravind, "A subcarrier allocation algorithm for OFDMA using buffer and channel state information", Vehicular Technology Conference, 2005.VTC-2005-Fall.2005 IEEE 62nd, 2005, pp.622-625.
- [38] Jun Shen, Na Yi, An Liu and Haige Xiang, "Opportunistic scheduling for heterogeneous services in downlink OFDMA system," School of EECS, Peking University, Beijing, P.R.China, IEEE computer Society 2009, pp.260- 264.
- [39] T. Badarch and O. Bataa, "Incoming traffic modeling of heterogeneous Public Safety Network," 2013 15th Asia-Pacific Network Operations and Management Symposium (APNOMS), Hiroshima, Japan, 2013, pp. 1-6
- [40] Degambur, Lavanya-Nehan & Mungur, Avinash & Armoogum, Sheeba & Pudaruth, Sameerchand. (2021). Resource Allocation in 4G and 5G Networks: A Review. International Journal of Communication Networks and Information Security (IJCNIS). 13. 10.54039/ijenis.v13i3.5116.
- [41] 5G NR Deployment Scenarios or modes-NSA,SA,Homogeneous,Heterogeneous". rfwireless-world.com.
- [42] Junko Yoshida. "What's Behind 'Non-Standalone' 5G?". Eetimes.com. Retrieved 2018-11-13.
- [43] Ameigeiras, P., Wigard, J., & Mogensen, P. (Sep. 2004). Performance of the m-lwdf scheduling algorithm for streaming services in hsdpa. In IEEE Transactions on vehicular technology conference vol. 2, pp. 999–1003 Los Angeles, USA.
- [44] Choi, J.-G., & Bahk, S. (2007). Cell-throughput analysis of the proportional fair scheduler in the single-cell environment. IEEE Transactions on Vehicular Technology, 56(2), 766–778.
- [45] Sadiq, B., Madan, R., & Sampath, A. (2009). Downlink scheduling for multiclass traffic in LTE. EURASIP Journal on Wireless Communications and Networking, 2009, 1–18
- [46] Habaebi MH, Chebil J, Al-Sakkaf AG, Dahawi TH. Comparison between scheduling techniques in long term evolution. IIUM EJ. 2013;14(1)
- [47] R. Kausar, Y. Chen and K. K. Chai, "Service Specific Queue Sorting and Scheduling Algorithm for OFDMA-Based LTE-Advanced Networks," 2011 International Conference on Broadband and Wireless Computing, Communication and Applications, Barcelona, Spain, 2011, pp. 116-121, doi: 10.1109/BWCCA.2011.22.
- [48] T. Cui, F. Lu, V. Sethuraman, A. Goteti, S. P. N. Rao and P. Subrahmanya, "Throughput Optimization in High Speed Downlink Packet Access (HSDPA)," in IEEE Transactions on Wireless Communications, vol. 10, no. 2, pp. 474-483, February 2011, doi: 10.1109/TWC.2010.120610.091294.



NARANTUYA Vandantseren in 2002 graduated from the Mongolian University of Science and Technology majoring in telecommunication. Bachelor degree thesis: Investigation Asynchronous Transfer Mode Technology, Master degree thesis (M.Sc.) in 2004: 3G mobile communication technology research. Doctoral student of Mongolian University of Science and Technology. Works in Mongolian National Defense University. Research topic: The Development of Defense communication service based on Mobile broadband convergence network.



CHULUUNBANDI Naimannaran, in 1994 graduated from the Mongolian University of Science and Technology majoring in information engineering and electronics. Master degree thesis (M.Sc.) in 2004: Optimization of Rural Radio communications. PhD degree thesis in 2012: Graph theory based performance evaluation for Rural cellular network services. Associate Professor, Consulting engineer of Mongolia. Research direction: Mobile Broadband technologies (WiMAX, WiBro, Mobile IPTV, 4G LTE, 5G technology etc.).



TUYATSETSEG Badarch, in 1996 graduated from the MUST majoring in Telecommunications engineer. Master degree thesis (M.Sc.) in 2013 in Northeastern University, USA: The Bayesian mixture model for traffic classification in Computer Networks. PhD degree in 2015: Performance Analysis of Integrated information network. MBA from Kansas Park University in 2017; Associate Professor at the Department of Computer Science at SICT of MUST. Research interest: Bayesian models for Traffic classification, clustering model in big data, AI.



ERDENE BAYAR Lamjav, in 2010 graduated from the Mongolian University of Science and Technology, Bachelor degree thesis: — The Study of Artificial Line. Master degree thesis (M.Sc.) in 2011: The study of link level simulation for Wimax. PhD degree thesis in 2018: The Study of Resource allocation in LTE networks. He is an Associate Professor at Department of Communication at SICT of MUST. Research direction: Mobile Broadband, high speed integrated services technologies (WiMAX, WiBro, Mobile IPTV, 4G LTE, 5G technology etc.).



OTGONBAYAR Bataa, in 1978 graduated from Polytechnic Institute of Mongolia majoring in Radio communication engineer. Bachelor degree thesis: Master degree thesis (M.Sc.) in 1995: Some issues of speech synthesis. PhD degree thesis in 1996: Study of Mongolian speech synthesis and applying it in telecom techniques, in 2003, post Ph.D. program thesis: Optimal version of OFDM system frequency and timing offset. Professor, Consulting engineer of Mongolia. Research topic: Mobile Broadband, high speed integrated services technologies.