Priority-aware Radio Resource Scheduling for mMTC in 5G Networks – Balancing Efficiency and Fairness

Prashant Kumar Baheti and Ajay Khunteta

Poornima University, Jaipur, Rajasthan, India

https://doi.org/10.26636/jtit.2026.1.2336

Abstract — Efficient and fair resource allocation for massive machine-type communication remains a significant challenge in 5G New Radio networks due to the diverse quality of service requirements and dynamic traffic patterns. This paper proposes a priority-aware uplink scheduling (PAUS) algorithm that jointly considers channel quality, 5G QoS identifier, packet aging, and fairness in physical resource block allocation, while simultaneously mitigating starvation of low-priority user equipment. The algorithm utilizes a composite fitness function to implement binary integer optimization for uplink scheduling, supported by heuristic resource assignment to ensure scalability. Simulation results demonstrate that the PAUS algorithm achieves an improved balance between throughput, resource utilization, delay, priority satisfaction, and fairness compared to baseline schedulers with polynomial-time complexity.

Keywords — 5G networks, efficiency, fairness, priority-aware scheduling, quality of service, radio resource scheduling

1. Introduction

The advent of 5G New Radio networks has fundamentally transformed wireless communications [1]. The International Telecommunication Union radiocommunication sector (ITU-R) introduced its 5G vision in 2015 through ITU-R recommendation M.2083, emphasizing that 5G NR would expand beyond enhanced mobile broadband (eMBB) to include massive machine-type communications (mMTC) and ultra-reliable low-latency communications (URLLC), each with stringent and diverse requirements [2].

According to [3], mMTC is characterized by low cost and low complexity of user equipment (UE), offering extended battery life, sporadic small data transmissions and relaxed latency constraints compared to URLLC applications. Different mMTC applications have different priority and delay tolerance characteristics, creating heterogeneous service requirements that pose unique challenges to radio resource allocation. 5G NR networks must ensure high throughput, massive connectivity, and optimal support for latency-tolerant and mission-critical applications [4].

Machine-type communication (MTC) represents a paradigm shift from traditional wireless communications. The integration of MTC into 5G NR networks introduces fundamental changes in traffic patterns [5], resource utilization, and scheduling requirements compared to previous generations. Most MTC traffic is uplink driven, with UEs periodically or sporadically transmitting data, status updates, or alarm notifications.

The 5G NR uplink framework introduces several key enhancements capable of addressing mMTC requirements. Flexible system-supporting subcarrier spacings equaling from 15 to 240 kHz may enable the allocation of resources for distinct mMTC service classes. Additionally, the bandwidth part (BWP) configuration allows for dynamic spectrum allocation to meet specific mMTC application requirements [6].

Despite these improvements, mMTC uplink scheduling in 5G NR faces several challenges, for example extreme heterogeneity in priority requirements, bursty traffic patterns with varying delay sensitivities, energy constrained UEs, massive connectivity scenarios that require scalable scheduling algorithms for high number of UEs, and the necessity to maintain long-term fairness to balance fair resource distribution with adherence to application priorities [7].

Traditional uplink scheduling algorithms such as proportional fair (PF) best channel quality (BestCQI) were optimized primarily for LTE communication, with relatively uniform traffic patterns and QoS requirements. In contrast, mMTC scenarios introduce diverse QoS requirements and varying UE capabilities across service classes, necessitating priority-aware scheduling mechanisms with dynamic physical resource block (PRB) assignment [8].

The scheduling problem also intensifies under high UE densities, due to resource shortages and the need to simultaneously support grant-based and grant-free transmission modes. Existing scheduling algorithms often fail to optimally integrate all relevant QoS indicators into effective PRB allocation decisions, and their performance is degraded under massive connectivity and heterogeneous service requirements [9].

Therefore, 5G NR networks introduce complexity to uplink PRB scheduling, especially since they converge diverse service requirements, flexible frame structures, and massive UE quality. Efficient and fair scheduling solutions are essential to realize the potential of 5G NR for mMTC applications – from low-priority environmental sensors to mission-critical automation systems.

Tab. 1. Comparison of key papers on priority-aware uplink scheduling for mMTC in 5G.

Ref.	Methodology	Traffic types	Contribution	Results
[24]	Heuristic and DRL	Heterogeneous mMTC	Adaptive GF resource partitioning	Improved fairness, lower packet drop rate, better resource utilization
[25]	Simulation-based	Smart city mMTC	Fairness and priority-based scheduling	Outperforms PF and BestCQI in terms of fairness and utilization
[26]	CTMC queuing model	eMBB, URLLC, mMTC	Priority-based allocation in C-RAN slicing	Higher resource utilization, reduced forced termination for mMTC
[27]	Dynamic priority assignment	Multi-RAT MTC	Two-stage dynamic priority scheduling	20 – 30% improvement in outage/success probability
[28]	Mathematical modeling	eMBB, URLLC, mMTC	Mixed reservation/priority RAN slicing	Up to 95% resource utilization, improved isolation

This work addresses these challenges by proposing a priority-aware uplink radio resource scheduling framework for mMTC scenarios that optimally balances efficiency and fairness, adheres to 5G NR standards, and addresses the unique QoS requirements of dense mMTC deployments.

1.1. Literature Review

The author of [10] proposes a 5G scheduler to minimize the long-term average age of information (AoI) for dense urban mMTC scenarios in which UEs frequently connect and disconnect, and adapts per-slot scheduling decisions with a priority algorithm to meet delay needs and reduce the average AoI by 10%, the deadline violation rate by 40%, and the consecutive violation rate by 20%. Similarly, [11] proposes a service-based M2M scheduling approach that integrates UE priority, service type, and channel quality to optimize throughput and resource utilization.

The studies also focus on deep reinforcement learning (DRL)-based research. Dynamic DRL-based slicing selects optimal slices under fluctuating demands while minimizing violations of service level agreements and allowing virtual network function migration for resource contention scenarios [12]. DRL-based resource slicing for multiservice coexistence provides near-optimal decoding and sum-rate performance according to [13]. In vehicular networks, slicing with QoS support prioritizes emergency traffic and maintains optimal throughput and reliability under road-specific conditions [14].

Paper [15] focused on traffic modeling for mMTC and revealed that despite growing UE numbers, aggregated inter-arrival times converge to a Poisson process, suggesting predictable scheduling opportunities even in massive deployments. For mMTC based on DECT-2020 NR, the MAC-layer design for power control and relay selection affects the packet delivery success, improving it by up to 10% [16].

The authors [17] propose a priority-enabled grant-free (GF) transmission access scheme that dynamically allocates slots within each 5G NR subframe, giving first access while assigning remaining slots to low-priority traffic, and develop a two-dimensional Markov chain model that accounts for both types of traffic. Furthermore, paper [18] proposes a scheduling mechanism to optimize resource allocation by addressing the trade-off between throughput, energy efficiency, and fairness through dynamic allocation of radio resources.

In [19], a fairness-aware uplink resource allocation scheme and an optimization problem are presented to ensure QoS satisfaction, using a demand-oriented greedy algorithm for PRB assignment and a bisection method for power allocation in UE, achieving performance improvements of up to 44.3% in fairness and 19.17% in QoS satisfaction. Furthermore, [20] proposes the min–max rate (MMR) algorithm to improve throughput and fairness compared to standard scheduling algorithms. In the same context, article [21] proposes an uplink scheduler to address spatial and frequency domain coupling, while achieving superior performance with polynomial-time complexity.

The authors of [22] address the PRB allocation problem as a mixed integer linear program (MILP), which utilizes mixed numerology, latency, and throughput requirements, multiple slices per UE, and internumerology interference, and propose a solution based on DRL. Also, article [23] proposes two schedulers based on the shortest job first (SJF) principle for minimum latency, and another for the joint optimization of energy efficiency (EE) and latency for complex multicellular scenarios.

Table 1 presents a comparison of key studies focusing on priority-based uplink scheduling for mMTC in 5G networks. The review demonstrates that priority-aware uplink scheduling is essential for supporting diverse mMTC requirements. Although these schemes are capable of reducing latency as well as improving resource utilization and fairness, the trade-off between prioritizing critical traffic and maintaining fairness for all UEs remains a challenge, particularly in congested scenarios with mixed traffic types and delay constraints.

1.2. Research Motivation and Contribution

The increase in the number of mMTC UEs requires optimal resource scheduling solutions capable of addressing diverse QoS constraints, latency budgets, and dense deployment challenges. In 5G NR, differentiated services enabled by 5QI require scheduling frameworks that ensure fast transmission

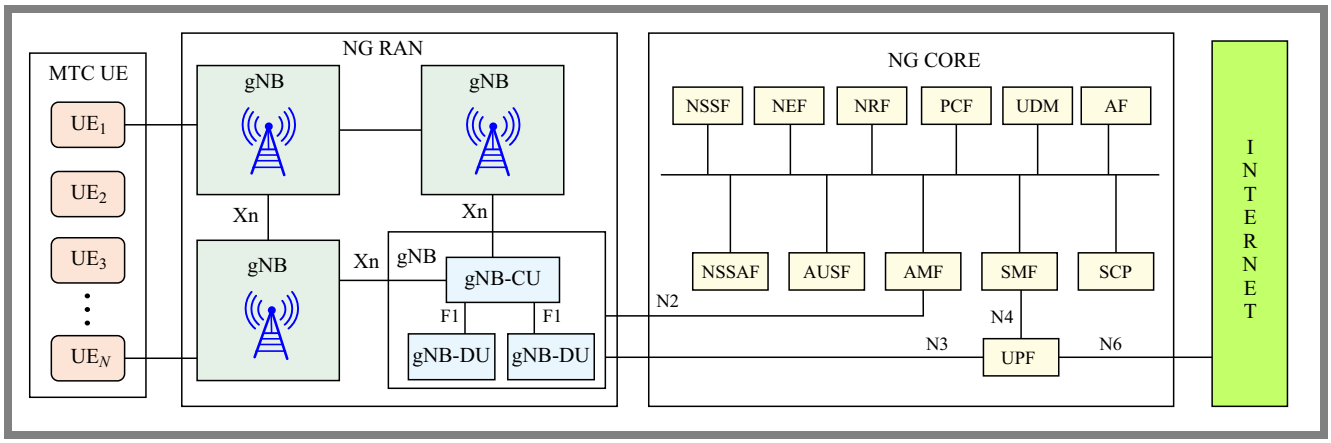

Fig. 1. 5G NR architecture components and interfaces [29].

for priority traffic, while maintaining fair resource sharing among heterogeneous UEs. Overcoming traditional scheduling limitations, such as the starvation of low-priority UEs and inefficient PRB utilization, requires optimal scheduling algorithms that explicitly balance efficiency with fairness.

The contributions of this work are as follows:

- It proposes a dynamic fitness function that integrates instantaneous channel quality, 5QI priorities, packet aging, and fairness adjustment in uplink scheduling decisions.
- It formulates the scheduling challenge as a constrained binary integer program, solved with greedy heuristics, suitable for real-time implementation with polynomial complexity.
- It implements comprehensive simulation experiments using the 5G NR system simulation library, benchmarking against BestCQI, priority-only, packet scheduling, and proportional fair algorithms. The outcomes are demonstrated with the use of comprehensive performance metrics such as throughput, resource utilization, fairness, priority satisfaction, and delay violation rate.

The remainder of this paper is organized as follows. Section 2 provides a background of the 5G network. Section 3 describes the proposed priority-aware scheduling algorithm. Section 4 presents the performance evaluation methodology. Section 5 discusses the results, evaluation, and analysis. Finally, Section 6 concludes the article.

2. 5G Network Background

This section provides an overview of the 5G NR architecture, the QoS framework, and the fundamentals of the radio resource grid. 5G NR networks can be deployed based on non-standalone (NSA) or standalone (SA) architectures. The SA architecture uses the complete 5G core (5GC), while the NSA extends the existing LTE infrastructure.

As shown in Fig. 1, the essential components are the UE, the radio access network (RAN), and the service-based 5GC. The main element of RAN is a set of gNB (5G node B) or (next generation node B) units, i.e. the base station and is interconnected by the Xn interface. A gNB can be subdivided into a central gNB unit (gNB-CU) and gNB distributed units (gNB-DU), connected by the F1 interface, supporting FDD and TDD modes. 5GC comprises modular network functions (NF) such as:

- user plane function (UPF) for data transport and data network (DN) for external connectivity,
- control plane for access and mobility management function (AMF), session management function (SMF), application function (AF), unified data management (UDM), policy control function (PCF), network repository function (NRF) and network slice selection function (NSSF).

The separation of control and user planes, which supports flexible user/data anchoring and virtualization of network functions, distinguishes 5G from LTE [30].

2.1. 5G QoS Framework

The 5G NR QoS framework supports diverse service requirements using a flow-based model, where each service flow is mapped to a data radio bearer (DRB) and is identified by a 5QI. The 5QI ensures differentiation by defined values ranging from 1 to 86 [31]. Each 5QI value corresponds to a predefined set of QoS parameters, such as:

- resource type: guaranteed bit rate (GBR), delay critical GBR, and non-GBR,
- priority: small priority value indicates higher priority,
- packet delay budget (PDB): maximum allowable delay D_{max} [ms],
- packet error rate (PER): maximum error rate,
- default maximum data burst volume: for delayed critical GBR services.

2.2. Priority and Delay Management

Let $r_{i,p}$ and $r_{j,q}$ represent PRB allocations for UE i and j with priority p and q, respectively. A priority-based QoS violation occurs if:

$$\mathcal{V}_{p,i}(t) = \begin{cases} 1 & \text{if } r_{i,p} = \emptyset \text{ and } r_{j,q} \neq \emptyset \ \ (p < q) \\ 0 & \text{otherwise} \end{cases} . \quad (1)$$

Each flow delay is managed with its PDB, D_i^{max}. If the network access delay D_i^n for UE i exceeds D_i^{max}, a delay violation occurs, thus:

$$D_i^n \leqslant D_i^{\max}, \quad \forall i \in \mathcal{U}(t)\,. \tag{2}$$

A QoS violation for delay-bound MTC occurs if:

$$\mathcal{V}_{d,i}(t) = \begin{cases} 1, & \text{if } D_i^n > D_i^{\max} \\ 0, & \text{otherwise} \end{cases}. \tag{3}$$

It ensures that the scheduler satisfies the priority and delay requirements for each QoS flow [32].

2.3. 5G Frame Structure and Resource Grid

5G NR uses flexible numerology μ ($\mu = 0, 1, 2, 3, 4$) to define the subcarrier spacing (SCS) $\Delta f = 2^\mu \times 15$ kHz, impacting slot duration and radio frame structure. Each radio frame spans 10 ms, comprising 10 subframes (1 ms each). Unlike LTE, which uses a fixed 15 kHz SCS, NR supports 15, 30, 60, 120, and 240 kHz SCS, allowing slot durations as low as 0.0625 ms for low-latency services [33].

Numerologies enable scalable, low-latency, or high-bandwidth scheduling, with slot durations decreasing as SCS increases. Each slot contains 14 orthogonal frequency-division multiplexing (OFDM) symbols (normal cyclic prefix). Table 2 presents the supported flexible transmission numerologies in 5G. As shown in Fig. 2, the resource grid represents the allocation of time-frequency resources as a two-dimensional grid with the frequency axis representing subcarriers ($N_{grid,x}^{size,\mu} \times N_{sc}^{PRB}$) and the time axis representing OFDM symbols ($N_{symb}^{slot} \times 2^\mu$), where PRB is the collection of 12 subcarriers in the frequency domain for one slot duration and resource element (RE) is the smallest resource unit representing one subcarrier for one OFDM symbol duration.

2.4. 5G Radio Resource Scheduling Framework

The scheduler, a key function of the gNB, allocates time-frequency resources to UEs for uplink and downlink communication. It is located within the medium access control (MAC) layer, but its operation is coordinated with the radio link control (RLC) and the physical layer (PHY) to enable efficient resource mapping [36].

The scheduling framework is driven by a set of metrics and reports transmitted by the UE. The gNB is based on a combination of UE-reported and gNB-measured information to efficiently allocate resources [37].

Scheduling metrics include uplink channel state measurement, buffer status report (BSR) and scheduling request, traffic pattern, power headroom report, logical channel configuration, and hybrid automatic repeat request (HARQ) [38].

PRB allocations in NR can be either contiguous or non-contiguous for a given UE. NR defines multiple types of the resource allocation process [39]:

- type 0 (resource block group, RBG-based, allowing non-contiguous allocation across the BWP,

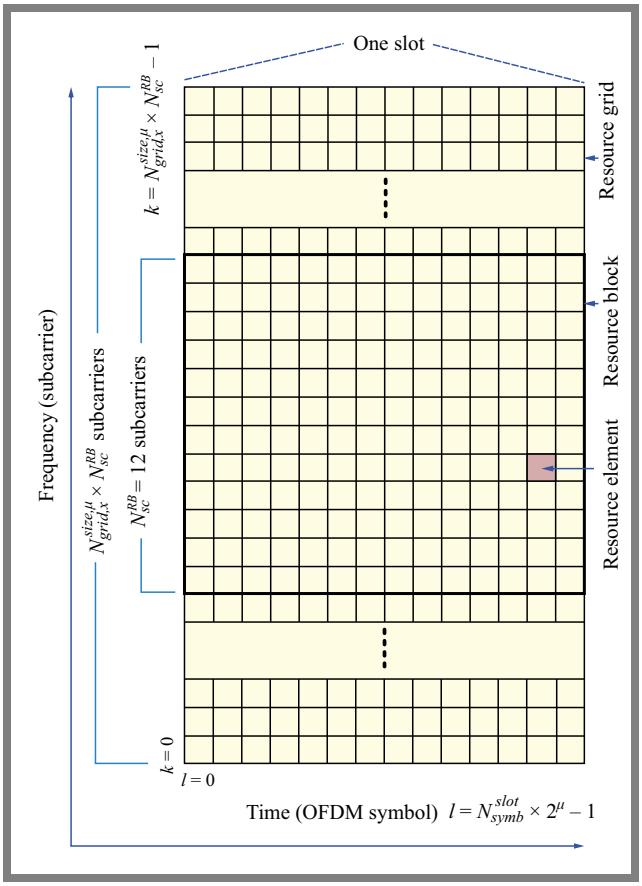

Fig. 2. Resource grid in 5G NR: subcarriers, OFDM symbols, PRB, and RE [35]

- type 1 (RB-based, enabling contiguous resource assignments for each UE).

The proposed scheduler considers service classes and dynamically assigns PRBs for sporadic, bursty mMTC traffic. The BSR, PDB, and throughput requirements are implemented in resource assignment.

3. Priority-aware Uplink Scheduling

3.1. Network Model

Consider a 5G NR cell with a single gNB serving a set $\mathcal{U} = \{1, 2, \ldots, N\}$ of mMTC UEs that are randomly distributed within the cell coverage area. At any scheduling interval t, let $\mathcal{X}_t \subseteq \mathcal{U}$ represent the subset of eligible UEs with buffered data. Each UE $i \in \mathcal{X}_t$ has a pending uplink grant request, with a resource requirement of r_i^t PRBs. Scheduling decisions are made at each transmission time interval (TTI) of the duration T_{TTI}. The total number of PRBs available in the cell is M [40].

3.2. Channel Model

The wireless channel between UE i and gNB is characterized by the channel quality indicator (CQI), $c_i^t \in \{0, 1, 2, \ldots, 15\}$. CQI is derived from signal-to-interference-plus-noise ratio (SINR) measurements using an effective SINR mapping

Tab. 2. 5G NR-supported flexible transmission numerologies [34].

μ	SCS [kHz]	FR1	FR2	PRB BW	Slots/frame	Slot duration	Symbol duration
0	15	√	×	180 kHz	10	1000 µs	66.67 µs
1	30	√	×	360 kHz	20	500 µs	33.33 µs
2	60	√	√	720 kHz	40	250 µs	16.67 µs
3	120	×	√	1.44 MHz	80	125 µs	8.33 µs
4	240	×	√	2.88 MHz	160	62.5 µs	4.17 µs

procedure:

$$\mathrm{SINR}_{\mathrm{eff},i}^{t} = \beta_1 \mathcal{I}^{-1}\left(\frac{1}{N_s}\sum_{k=1}^{N_s}\mathcal{I}\frac{\mathrm{SINR}_{i,k}^{t}}{\beta_2}\right), \tag{4}$$

where N_s is the number of SINR samples, $\mathcal{I}(\cdot)$ is the mutual information function and β_1, β_2 are parameters specific to the modulation and coding scheme (MCS).

Furthermore, [41] specifies a mapping table that relates the resulting SINR to the CQI and the transport block size (TBS) index, and we use it.

3.3. Dynamic Fitness Function

The proposed priority-aware uplink scheduling (PAUS) algorithm employs a dynamic fitness function to maximize the utility of the system by considering channel quality, QoS priority, packet aging, and long-term fairness. For each UE $i \in \mathcal{X}_t$ at scheduling time t, the fitness function is defined as:

$$F_i^t = \omega_i^t + \alpha_i^t \cdot \left(F_i^{t-1} + \xi_i^t\right), \tag{5}$$

where ω_i^t is the channel-aware priority weight, α_i^t is a temporal decay factor, F_i^{t-1} is the historical fitness value from the previous TTI ($F_i^0 = 0$) and ξ_i^t is a fairness adjustment term. The fitness value acts as a dynamic scheduling weight that evolves over time.

To balance efficiency with QoS requirements, we define a priority weight that combines channel conditions and static QoS priority. We assume that each UE i has a static priority $p_i \in \{10, 20, \ldots, P\}$, where lower values indicate a higher priority (i.e., $p_i = 10$ is the highest priority). We normalize CQI to [0,1], as $c_{i,\mathrm{norm}}^t = c_i^t/15$. The channel-aware priority weight is formulated as:

$$\omega_i^t = (c_{i,\mathrm{norm}}^t)^{\gamma} \cdot \left(1 + \frac{P - p_i}{P}\right), \tag{6}$$

where $\gamma \in [0.5, 2]$ is a scaling parameter that controls the influence of channel quality and P is the lowest priority index. This ensures that a higher CQI combined with a higher QoS priority increases the chances of UE scheduling.

To prevent excessive delays for buffered packets and ensure timely delivery for delay-sensitive applications while maintaining system stability, a temporal decay factor is introduced:

$$\alpha_i^t = \begin{cases} e^{-\beta\cdot(t - t_i^{\mathrm{last}})} & \text{if } t - t_i^{\mathrm{last}} < D_i^{\max} - \epsilon \\ 0 & \text{otherwise} \end{cases}, \tag{7}$$

where t_i^{last} is the last scheduling time for UE i, D_i is the PDB for UE i, $\beta \in [0.01, 1]$ is the decay rate parameter, $0 \leqslant \epsilon \leqslant 0.2 \times D_i$, is a tolerance threshold, and $\alpha_i^t \in [0, 1]$. The decay mechanism ensures that UEs are scheduled before their packet deadline while resetting at the edge of the deadline to mitigate QoS delay violations and enforces soft deadline scheduling.

To mitigate the starvation of low-priority UEs, we incorporate a fairness adjustment term that provides a gradually increasing boost to UEs that have not been scheduled for a long time.

$$\xi_i^t = \frac{\eta}{1 + e^{-\lambda\cdot(t - t_i^{\mathrm{last}})}}, \quad 0 \leqslant \xi_i^t \leqslant \eta, \tag{8}$$

where $\eta \in [0.1, 2]$ controls the maximum fairness boost and $\lambda \in [0.05, 2]$ determines the steepness of the sigmoid function. This term implements a bounded increase in fitness score proportional to waiting time, ensuring that no UE experiences starvation regardless of its priority or channel conditions. Sigmoid-based term increases fitness for prolonged unscheduled UEs, but is upper bounded by η, preventing instability.

The specific default values and recommended ranges for γ, β, η, and λ were determined by the R-method [42], a technique to classify multiple weighted criteria or parameters.

3.4. Optimization Problem Formulation

The dynamic uplink scheduling problem is formulated as a constrained binary integer optimization problem aimed at maximizing the aggregate fitness of scheduled UEs while respecting resource and operational constraints. The decision variables are the UE scheduling vector x_t and the PRB assignment matrix $\mathbf{s}_t$:

$$\max_{x_t, \mathbf{s}_t} \sum_{i \in \mathcal{X}_t} x_i^t \cdot F_i^t, \tag{9}$$

subject to

$$\sum_{i \in \mathcal{X}_t} x_i^t \cdot r_i^t \leqslant M, \tag{10}$$

$$x_i^t \in \{0, 1\}, \quad \forall i \in \mathcal{X}_t, \tag{11}$$

$$\sum_{i \in \mathcal{X}_t} s_{i,j}^t \leqslant 1, \quad \forall j \in \{1, \ldots, M\}, \tag{12}$$

$$\sum_{j=1}^{M} s_{i,j}^t = r_i^t \cdot x_i^t, \quad \forall i \in \mathcal{X}_t, \tag{13}$$

$$s_{i,j}^t \in \{0, 1\}, \quad \forall i \in \mathcal{X}_t, j \in \{1, \ldots, M\}. \tag{14}$$

The optimization problem forms Eqs. (9) – (14) and determines the set of UEs to be scheduled and the corresponding PRB-level allocation within each TTI. Constraint (10) enforces the PRB limitation of the uplink bandwidth by ensuring that the aggregate PRB requirement of all selected UEs does not exceed the total number of PRBs available. The binary constraint (11) specifies whether UE i is selected for transmission and manages the scheduling decision, while (12) enforces PRB exclusivity by ensuring that each PRB is assigned to no more than one UE, reflecting the orthogonality requirement of NR uplink transmissions. Completeness constraint (13) imposes an all-or-nothing allocation rule, ensuring that a UE receives either the full set of PRBs it requests or nothing, thereby avoiding partial transport block formation and simplifying HARQ operations [43].

Finally, constraint (14) ensures the binary PRB assignment and manages resource mapping. Therefore, the formulation captures the discrete, non-linear, and coupled nature of uplink scheduling in 5G NR, making the problem NP hard and motivating the use of heuristic algorithms such as the proposed PAUS Algorithm 1.

Because the uplink scheduling problem is formulated in Eqs. (9) – (14) is NP-hard, a greedy heuristic is adopted to obtain an efficient real-time solution. The algorithm computes the fitness value for each eligible UE and iteratively selects the UEs in descending order of fitness until all PRBs are exhausted. This procedure ensures that the scheduler does not forget about UEs that have been waiting, making it inherently fair while still being efficient. The uplink scheduling workflow is illustrated in Fig. 3.

The computational complexity of the PAUS algorithm is dominated by the UE fitness sorting step. Calculating the fitness function for all eligible UEs requires $O(|\mathcal{X}_t|)$ operations. The sorting step has the complexity of $O(|\mathcal{X}_t| \log |\mathcal{X}_t|)$. The greedy resource allocation loop executes at most $O(|\mathcal{X}_t|)$ iterations. Therefore, the overall time complexity is $O(|\mathcal{X}_t| \log |\mathcal{X}_t|)$, which is polynomial and feasible for real-time scheduling, where $|\mathcal{X}_t|$ may range from tens to thousands of UEs per TTI.

Algorithm 1 PAUS algorithm for dynamic scheduling

Input: $\mathcal{X}_t$, $\{p_i, c_i^t, r_i^t, D_i, t_i^{\text{last}}, F_i^{t-1}\}_{i\in\mathcal{X}_t}$, $M, \gamma, \beta, \eta, \lambda, \epsilon$
Output: scheduled UE set $\mathcal{S}_t$, updated fitness $\{F_i^t\}_{i\in\mathcal{X}_t}$, PRB allocation map $\mathbf{s}_t$
Initialize: $\mathcal{S}_t = \emptyset$, $M_{\text{rem}} = M$
for each $i \in \mathcal{X}_t$ **do**
    Compute ω_i^t using Eq. (6)
    Compute α_i^t using Eq. (7)
    Compute ξ_i^t using Eq. (8)
    Compute F_i^t using Eq. (5)
end for
Sort UEs in $\mathcal{X}_t$ in descending order of F_i^t values
for each UE i in sorted order **do**
    if $r_i^t \leqslant M_{\text{rem}}$ **then**
        $\mathcal{S}_t = \mathcal{S}_t \cup \{i\}$
        $M_{\text{rem}} = M_{\text{rem}} - r_i^t$
        Allocate r_i^t PRBs to UE i
        Set $s_{i,j}^t = 1$ for all allocated PRBs j
        Update $t_i^{\text{last}} = t$
    end if
end for
Return $\mathcal{S}_t$, $\{F_i^t\}_{i\in\mathcal{X}_t}$ and PRB allocation map $\mathbf{s}_t$

4. Performance Evaluation

4.1. Simulation Setup

We consider a network scenario with a coverage radius of 1 km, where mMTC UEs are randomly distributed within a single 5G NR macro cell, and gNB is positioned at the cell's center. UEs communicate directly with the gNB without requiring dedicated gateways. UEs with the same 5QI priorities are grouped for traffic differentiation.

The simulation is implemented using the Matlab 5G NR system simulator. This framework provides comprehensive simulation capabilities, incorporating the PHY, MAC, and RLC layers, as well as configurable traffic generation. UEs remain stationary during each simulation run to isolate scheduler performance from mobility-based channel variations.

For PHY modeling, we utilize the integrated PHY abstraction in the 5G Toolbox, which implements link-to-system mapping for performance evaluation. The channel model is the 3GPP TR 38.901 urban macro (UM) scenario [44]. RLC entities operate in an unacknowledged mode to reduce protocol overhead, a suitable choice for latency-sensitive and small-payload mMTC traffic.

Scheduling decisions are executed in the MAC layer of the gNB, which supports the integration of custom uplink schedulers. The proposed PAUS scheduler utilizes a composite metric that incorporates instantaneous CQI feedback, 5QI priority, packet aging, and long-term fairness. The key simulation parameters are summarized in Tab. 3. A duration of

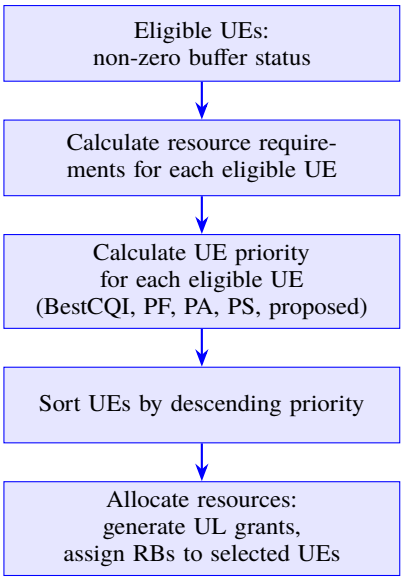

Fig. 3. Particle swarm optimization flow chart.

Tab. 3. Simulation parameters.

Parameter	Value
Channel bandwidth	5 MHz
Number of PRBs	25
Duplex mode	FDD
Number of gNBs	1
Numerology (μ)	0 (15 kHz SCS)
Subcarriers per PRB	12
OFDM symbols per slot	14
Slots per subframe	1
Subframes per frame	10
Frame duration	10 ms
Slot duration	1 ms
Channel model	Urban macro (UM)
Number of UEs	100, 200, . . . , 500
UE positions	Random
UE height	1.5 m
Simulation duration	500 frames

500 frames provides sufficient time to observe steady-state scheduler behavior under varying mMTC loads while keeping simulation runtime practical.

4.2. Application Scenario

The evaluation considers a smart city deployment characterized by high-density machine-type connectivity. mMTC UEs, such as smart meters, environmental sensors, public asset monitors, and traffic control are utilized. Their communication pattern consists mainly of sporadic small payload uplink transmissions. The primary challenge for the 5GC and NR RAN is to efficiently support massive connectivity while minimizing signaling load and satisfying heterogeneous QoS requirements. The traffic distribution across 5QI classes is presented in Tab. 4 [45].

4.3. Baseline Algorithms

To evaluate PAUS performance, we implement four benchmark algorithms for comparison purposes:

- Best CQI (BestCQI) scheduler. It allocates PRBs to UEs with the highest instantaneous channel quality, maximizing throughput but ignoring fairness, calculated from channel-dependent MCS, transmission rank, and available REs.
- Priority algorithm (PA). It schedules UEs strictly based on static 5QI priority, independently of channel conditions and fairness.
- Packet scheduling (PS). This algorithm [46] combines QoS requirements with the fairness mechanism, but fairness deteriorates drastically with network load.
- Proportional fair (PF) scheduler. It balance throughput and fairness by prioritizing UEs with high instantaneous-to-average rate ratios, using an exponential moving average.

4.4. Performance Metrics

The following key performance indicators are used to assess the efficiency of uplink scheduling algorithms.

The average cell throughput R_{cell} measures the mean data rate successfully delivered from all scheduled UEs to the gNB and is defined as:

$$R_{\text{cell}}\,[\text{Mbps}] = \frac{\left(\sum_{t=1}^{T_{\text{slot}}}\sum_{i=1}^{N} R_i^t\right)\times 8}{T_{\text{sim}}\times 10^6}\,, \tag{15}$$

where R_i^t is the data [bytes] received from UE i in slot t, T_{slot} is the total number of slots, T_{sim} ($\text{NumFrames}\times 0.01$ s/frame) is the total simulation time in seconds, and N is the total number of active UE.

Resource utilization ρ quantifies the ratio of PRBs allocated to the total PRBs available and is defined as:

$$\rho = \frac{1}{T_{\text{slot}}}\sum_{t=1}^{T_{\text{slot}}}\frac{\sum_{i=1}^{N} r_i^t}{M}, \qquad 0\leqslant\rho\leqslant 1\,, \tag{16}$$

where r_i^t is the number of PRBs allocated to UE i in the slot t and M is the total number of PRBs available per slot.

Instantaneous resource utilization is defined as:

$$\rho(t) = \frac{\sum_{i=1}^{N} r_i^t}{M}\,. \tag{17}$$

The fairness index $\mathcal{F}$ evaluates the equitable distribution of radio resources among competing UEs, using Jain's fairness index is:

$$\mathcal{F} = \frac{\left(\sum_{i=1}^{N} R_i\right)^2}{N\cdot\sum_{i=1}^{N} R_i^2}, \qquad 0\leqslant\mathcal{F}\leqslant 1\,, \tag{18}$$

where each R_i is the throughput for UE i:

$$R_i = \frac{\left(\sum_{t=1}^{T_{\text{slot}}} R_i^t\right)\times 8}{T_{\text{sim}}\times 10^6}\,. \tag{19}$$

5QI priority satisfaction evaluates the proportion of scheduling events where the scheduler fails to satisfy QoS priority differentiation among traffic classes. Lower values indicate better 5QI priority satisfaction and improved QoS differentiation.

The delay violation rate represents the average number of events when the packets exceed their PDB before successful transmission. A lower delay violation rate indicates better delay performance.

5. Results and Discussion

This section analyzes the experimental results obtained from extensive simulations for a range of mMTC network sizes

Tab. 4. mMTC traffic parameters for the cell radius r = 1000 m.

Priority	5QI	D_i^{max} [ms]	Packets/min	Traffic type	UE [%]	Example service
10	5	100	5	Event update	12	Emergency notification service
20	1	100	10	Event update	7	Intrusion detection service
30	3	50	30	Periodic update	12	Critical monitoring service
40	2	150	30	Event/periodic update	12	Environment monitoring
50	4	300	40	Event/periodic update	19	Smart home service
60	6	300	60	Periodic update	19	Utility application service
70	7	100	60	Periodic update	19	Smart meters

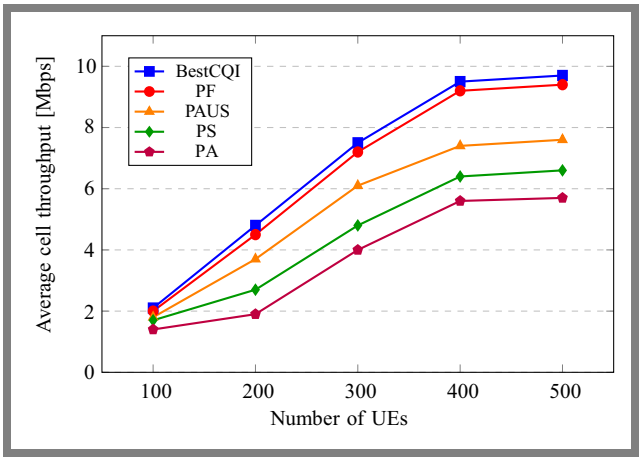

Fig. 4. Average cell throughput.

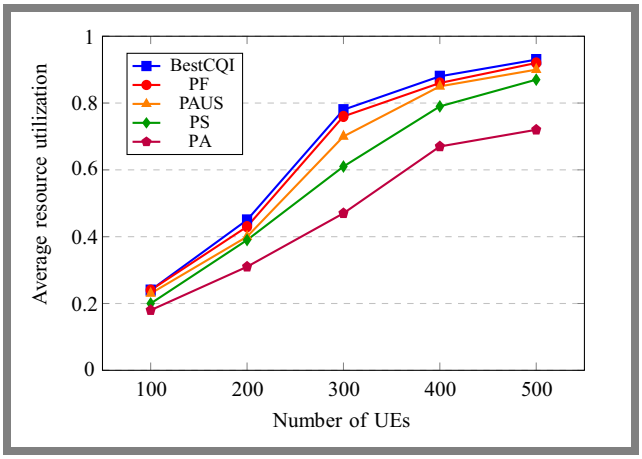

Fig. 5. Average resource utilization.

and load conditions. Each experiment is repeated for 50 independent runs. The reported results correspond to the sample mean across the runs.

5.1. Average Cell Throughput

Average cell throughput is defined in Eq. (15). Figure 4 compares the average cell throughput of the five schedulers under consideration as the number of UEs increases from 100 to 500. Channel-oriented schemes (BestCQI and PF) achieve the highest throughput because they prioritize instantaneous channel capacity. The PAUS scheduler achieves optimal throughput while improving fairness and priority satisfaction. Importantly, PAUS outperforms the PA and PS baselines across all network sizes by combining channel awareness with priority and aging information, aligning scalability and efficiency.

5.2. Resource Utilization

Resource utilization, defined in Eq. (16), quantifies the efficiency of PRB usage. Figure 5 shows that CQI-centric schedulers (BestCQI, PF) achieve the highest utilization, exceeding 90% under heavy load. PAUS achieves high utilization (up to 90%) by opportunistically exploiting channel instances while respecting priority constraints. The PA algorithm underperforms in utilization at high loads because it schedules according to static priority, often leaving spectral opportunities unused when high-priority UEs have poor channel quality, underlining the importance of channel awareness for spectral efficiency.

5.3. Fairness Index

The fairness index is defined in Eq. (18), where values close to 1 denote high fairness. Figure 6 illustrates that PAUS achieves higher fairness ≈ 0.84 than a purely channel-oriented scheduler ≈ 0.30 and static priority ≈ 0.46 at 500 UEs. PF exhibits intermediate fairness. The high fairness of PAUS is due to the sigmoid-based fairness boost ξ_i^t and the aging-aware decay factor, which together mitigate the starvation of low-priority or persistently poor-channel UEs, suitable for dense, heterogeneous IoT environments.

5.4. Priority Satisfaction

Figure 7 shows that the PA baseline delivers the lowest violation rate (highest 5QI satisfaction), but at a cost to throughput and fairness. PAUS achieves a balanced trade-off with 0.23 for 500 UE, lower than PF and BestCQI, while maintaining optimal throughput and fairness than PA. PF and BestCQI prioritize instantaneous channel conditions, accumulating more 5QI priority violations. This shows that PAUS effectively balances QoS differentiation with minimum sacrifice of efficiency.

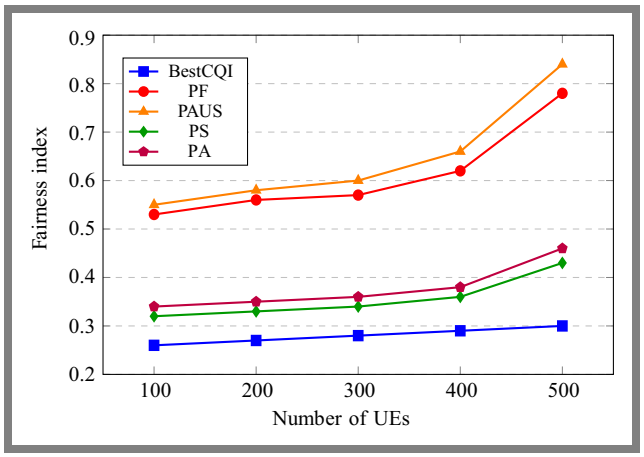

Fig. 6. Fairness index.

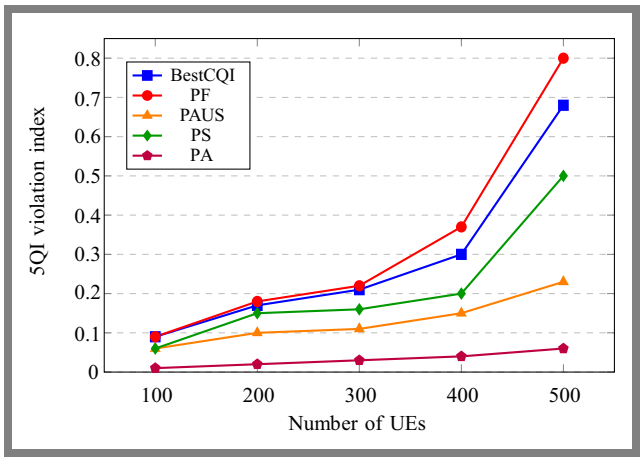

Fig. 7. 5QI priority satisfaction.

5.5. Delay Violation Rate

The delay violation rate is a critical mMTC reliability indicator. The results shown in Fig. 8 illustrate that PAUS yields the lowest average delay violation, ≈ 8.5 for 500 UE, due to the deadline-aware temporal decay α_i^t and the fairness boost that prioritizes aged packets, and is lower than all other schemes. PF and BestCQI, which ignore deadline information, incur substantially higher delay violations under heavy loads. The findings suggest the suitability of PAUS for delay-sensitive mMTC traffic.

The results demonstrate that PAUS achieves a balance between efficiency and fairness/QoS differentiation in dense mMTC deployments. Although BestCQI and PF maximize throughput, they suffer from poor fairness and higher 5QI violations. In contrast, PA satisfy priority at the cost of efficiency. PAUS achieves near-PF throughput with substantially improved fairness and priority satisfaction, making it suitable for urban mMTC scenarios.

6. Conclusions

This paper presents a priority-aware uplink radio resource scheduling algorithm to address the challenge of balancing efficiency with optimal fairness in dense 5G mMTC deployments. The novelty of the proposed algorithm lies in its dynamic fitness function, which integrates a channel-aware priority weight to utilize channel conditions, a temporal decay mechanism to mitigate delay violations based on the PDB, and a sigmoid-based fairness adjustment to prevent starvation of low-priority or poor-channel UEs.

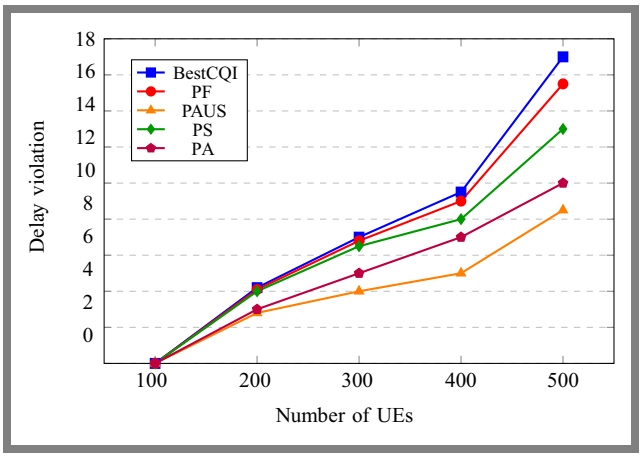

Fig. 8. 5QI priority satisfaction.

Extensive simulation results demonstrated that PAUS achieved near-optimal throughput and resource utilization, close to channel-aware schemes (PF, BestCQI). PAUS outperformed efficiency-focused schemes in QoS metrics and achieved the highest fairness and the lowest delay violation rate, which can manage delay-sensitive traffic. Furthermore, it maintained a low 5QI priority violation, demonstrating effective QoS differentiation with balanced spectral efficiency.

Polynomial-time complexity makes PAUS implementable within real-time 5G gNB MAC scheduler architectures and is suitable for supporting the large-scale, heterogeneous traffic requirements of urban smart city and industrial IoT applications.

References

[1] 3GPP, "5G; Study on New Radio (NR) Access Technology (3GPP TR 38.912 Version 18.0.0 Release 18)", Technical Report, 2024.

[2] J. Janković *et al.*, "Effects of Differentiated 5G Services on Computational and Radio Resource Allocation Performance", *IEEE Transactions on Network and Service Management*, vol. 18, pp. 2226–2241, 2021 (https://doi.org/10.1109/TNSM.2021.3060865).

[3] 3GPP, "5G; NR; NR and NG-RAN Overall Description; Stage-2 (3GPP TS 38.300 Version 17.0.0 Release 17)", Technical Specification, 2022.

[4] S.R. Pokhrel *et al.*, "Towards Enabling Critical mMTC: A Review of URLLC within mMTC", *IEEE Access*, vol. 8, pp. 131796–131813, 2020 (https://doi.org/10.1109/ACCESS.2020.3010271).

[5] S. Rezwan and W. Choi, "Priority-based Joint Resource Allocation with Deep Q-learning for Heterogeneous NOMA Systems", *IEEE Access*, vol. 9, pp. 41468–41481, 2021 (https://doi.org/10.1109/ACCESS.2021.3065314).

[6] M. Attaran, "The Impact of 5G on the Evolution of Intelligent Automation and Industry Digitization", *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, pp. 5977–5993, 2023 (https://doi.org/10.1007/s12652-020-02521-x).

[7] S. Hamdoun, A. Rachedi, and Y. Ghamri-Doudane, "Graph-based Radio Resource Sharing Schemes for MTC in D2D-based 5G Networks", *Mobile Networks and Applications*, vol. 25, pp. 1095–1113, 2020 (https://doi.org/10.1007/s11036-020-01527-1).

[8] P.K. Baheti and A. Khunteta, "QoS Aware Resource Scheduling in LTE Network for Smart City M2M Communication", *2021 IEEE International Conference on Technology, Research, and Innovation for Betterment of Society (TRIBES)*, Raipur, India, 2021 (https://doi.org/10.1109/TRIBES52498.2021.9751661).

[9] T.-S. Lee, C.-H. Yang, T.-Y. Kuo, and Y.-J. Wu, "Chapter 13 - Resource Allocation in Massive Machine-type Communications", in: *Resource Optimization in Wireless Communications*, pp. 317–347, 2025 (https://doi.org/10.1016/B978-0-44-330092-9.00019-7).

[10] B.-S. Kim, "A Priority-aware Dynamic Scheduling Algorithm for Ensuring Data Freshness in 5G Networks", *Future Generation Computer Systems*, vol. 163, art. no. 107542, 2025 (https://doi.org/10.1016/j.future.2024.107542).

[11] P.K. Baheti and A. Khunteta, "Service-aware Resource Scheduling for Heterogeneous M2M Communication in 5G Networks", *International Journal of Basic and Applied Sciences*, vol. 14, pp. 137–143, 2025 (https://doi.org/10.14419/saqrkp91).

[12] P. Vidhya, K. Subashini, R. Sathishkannan, and S. Gayathri, "Dynamic Network Slicing Based Resource Management and Service Aware Virtual Network Function (VNF) Migration in 5G Networks", *Computer Networks*, vol. 259, art. no. 111064, 2025 (https://doi.org/10.1016/j.comnet.2025.111064).

[13] S. Malta, P. Pinto, and M. Fernandez-Veiga, "Optimizing 5G Network Slicing with DRL: Balancing eMBB, URLLC, and mMTC with OMA, NOMA, and RSMA", *Journal of Network and Computer Applications*, vol. 234, art. no. 104068, 2025 (https://doi.org/10.1016/j.jnca.2024.104068).

[14] W. Hamdi, O. Dağdeviren, and H. Bulut, "QoS-aware Network Slicing and Resource Management for Internet of Vehicles in 5G Networks", *Ad Hoc Networks*, vol. 178, art. no. 103976, 2025 (https://doi.org/10.1016/j.adhoc.2025.103976).

[15] E. Goshi, F. Mehmeti, T.F. La Porta, and W. Kellerer, "Modeling and Analysis of mMTC Traffic in 5G Core Networks", *IEEE Transactions on Network and Service Management*, vol. 22, pp. 409–425, 2025 (https://doi.org/10.1109/TNSM.2024.3481240).

[16] A. Samuylov *et al.*, "Performance of MAC Layer Mechanisms in DECT-2020 NR mMTC Technology", *2024 IEEE 99th Vehicular Technology Conference (VTC2024-Spring)*, Singapore, Singapore, 2024 (https://doi.org/10.1109/VTC2024-Spring62846.2024.10683658).

[17] T.N. Weerasinghe, V. Casares-Giner, I.A.M. Balapuwaduge, and F.Y. Li, "Priority Enabled Grant-free Access with Dynamic Slot Allocation for Heterogeneous mMTC Traffic in 5G NR Networks", *IEEE Transactions on Communications*, vol. 69, pp. 3192–3206, 2021 (https://doi.org/10.1109/TCOMM.2021.3053990).

[18] M. Abdullah *et al.*, "Satellite Synergy: Navigating Resource Allocation and Energy Efficiency in IoT Networks", *Journal of Network and Computer Applications*, vol. 230, art. no. 103966, 2024 (https://doi.org/10.1016/j.jnca.2024.103966).

[19] Y.L. Lee, T.C. Chuah, J. Loo, and F. Ke, "Proportional-fair Uplink Resource Allocation with Statistical QoS Provisioning for RAN Slicing", *Physical Communication*, vol. 65, art. no. 102389, 2024 (https://doi.org/10.1016/j.phycom.2024.102389).

[20] M.O. Kabaou *et al.*, "Empowering Communication Networks with MMR Scheduler: A Novel Approach to Balancing User Throughput and Fairness", *Alexandria Engineering Journal*, vol. 76, pp. 641–649, 2023 (https://doi.org/10.1016/j.aej.2023.06.042).

[21] L. Zhang, A. Liu, and X. Chen, "A WMMSE-based Contiguous Resource Scheduling Algorithm for 5G-NR Uplink", IEEE Wireless Communications Letters, vol. 13, pp. 466–470, 2024 (https://doi.org/10.1109/LWC.2023.3332327).

[22] K. Boutiba, M. Bagaa, and A. Ksentini, "Optimal Radio Resource Management in 5G NR Featuring Network Slicing", *Computer Networks*, vol. 234, art. no. 109937, 2023 (https://doi.org/10.1016/j.comnet.2023.109937).

[23] O. Elgarhy *et al.*, "Energy Efficiency and Latency Optimization for IoT URLLC and mMTC Use Cases", *IEEE Access*, vol. 12, pp. 23132–23148, 2024 (https://doi.org/10.1109/ACCESS.2024.3364349).

[24] Y. Kaura, B. Lall, R.K. Mallik, and A. Singhal, "Adaptive Scheduling of Shared Grant-free Resources for Heterogeneous Massive Machine Type Communication in 5G and Beyond Networks", *IEEE Transactions on Network and Service Management*, vol. 22, pp. 1188–1204, 2025 (https://doi.org/10.1109/TNSM.2024.3493015).

[25] P.K. Baheti and A. Khunteta, "Priority-based Resource Scheduling for Smart City M2M Communication in 5G Networks", *2023 3rd International Conference on Mobile Networks and Wireless Communications (ICMNWC)*, Tumkur, India, 2023 (https://doi.org/10.1109/ICMNWC60182.2023.10435821).

[26] S.A. AlQahtani, "Cooperative-aware Radio Resource Allocation Scheme for 5G Network Slicing in Cloud Radio Access Networks", *Sensors*, vol. 23, art. no. 5111, 2023 (https://doi.org/10.3390/s23115111).

[27] W.U. Rehman *et al.*, "Improved Resource Allocation in 5G MTC Networks", *IEEE Access*, vol. 8, pp. 49187–49197, 2020 (https://doi.org/10.1109/ACCESS.2020.2974632).

[28] D. Ivanova *et al.*, "Mathematical Framework for Mixed Reservation- and Priority-based Traffic Coexistence in 5G NR Systems", *Mathematics*, vol. 11, art. no. 1046, 2023 (https://doi.org/10.3390/math11041046).

[29] S.B. Prathiba, K. Raja, R.V. Saiabirami, and G. Kannan, "An Energy-aware Tailored Resource Management for Cellular-based Zero-touch Deterministic Industrial M2M Networks", *IEEE Access*, vol. 12, pp. 33613–33627, 2024 (https://doi.org/10.1109/ACCESS.2024.3372417).

[30] M.R.M. Anfar and J. Mwangama, "Machine Learning-based Service Differentiation in the 5G Core Network", *2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, Jeju Island, South Korea, 2021 (https://doi.org/10.1109/ICAIIC51459.2021.9415263).

[31] 3GPP, "5G; System Architecture for the 5G System (5GS) (3GPP TS 23.501 Version 17.5.0 Release 17)", Technical Specification, 2022 (https://www.etsi.org/deliver/etsi_ts/123500_123599/123501/17.05.00_60/ts_123501v170500p.pdf).

[32] S.O. Oladejo and O.E. Falowo, "Latency-aware Dynamic Resource Allocation Scheme for Multi-tier 5G Network: A Network Slicing-multitenancy Scenario", *IEEE Access*, vol. 8, pp. 74834–74852, 2020 (https://doi.org/10.1109/ACCESS.2020.2988710).

[33] B. Agarwal, M.A. Togou, M. Marco, and G.-M. Muntean, "A Comprehensive Survey on Radio Resource Management in 5G HetNets: Current Solutions, Future Trends and Open Issues", *IEEE Communications Surveys & Tutorials*, vol. 24, pp. 2495–2534, 2022 (https://doi.org/10.1109/COMST.2022.3207967).

[34] A. Yazar and H. Arslan, "Flexible Multi-numerology Systems for 5G New Radio", *Journal of Mobile Multimedia*, vol. 14, pp. 367–394, 2018.

[35] E. Engin, I. Hokelek, A. Gorcin, and H.A. Cirpan, "A Pre-emptive Scheduling Mechanism for Service Assurance of Network Slicing in Next Generation Cellular Networks", *IEEE Access*, vol. 13, pp. 23297–23311, 2025 (https://doi.org/10.1109/ACCESS.2025.3536997).

[36] 3GPP, "5G; NR; Medium Access Control (MAC) Protocol Specification (3GPP TS 38.321 Version 16.1.0 Release 16)", Technical Specification, 2020.

[37] V. Kovtun, O. Kovtun, K. Grochla, and O. Yasniy, "The Quality of Service Assessment of eMBB and mMTC Traffic in a Clustered 5G Ecosystem of a Smart Factory", *Egyptian Informatics Journal*, vol. 29, art. no. 100598, 2025 (https://doi.org/10.1016/j.eij.2024.100598).

[38] T. Wang, "Energy-efficient Resource Allocation for UAV-aided Full-duplex OFDMA Wireless Powered IoT Communication Networks", *Journal of King Saud University - Computer and Information Sciences*, vol. 36, art. no. 102225, 2024 (https://doi.org/10.1016/j.jksuci.2024.102225).

[39] P.K. Korrai *et al.*, "Joint Power and Resource Block Allocation for Mixed-numerology-based 5G Downlink under Imperfect CSI", *IEEE Open Journal of the Communications Society*, vol. 1, pp. 1583–1601, 2020 (https://doi.org/10.1109/OJCOMS.2020.3029553).

[40] 3GPP, "5G; NR; Physical Channels and Modulation (3GPP TS 38.211 Version 16.2.0 Release 16)", Technical Specification, 2020.

[41] 3GPP, "5G; NR; Physical Layer Procedures for Data (3GPP TS 38.214 Version 16.2.0 Release 16)", Technical Specification, 2020.

[42] R.V. Rao and J. Lakshmi, "R-method: A Simple Ranking Method for Multi-attribute Decision-making in the Industrial Environment", *Journal of Project Management*, vol. 6, pp. 223–230, 2021 (https://doi.org/10.5267/j.jpm.2021.5.001).

[43] 3GPP, "5G; NR; Physical Layer Procedures for Control (3GPP TS 38.213 Version 17.1.0 Release 17)", Technical Specification, 2022.

[44] 3GPP, "5G; Study on Channel Model for Frequencies From 0.5 to 100 GHz (3GPP TR 38.901 Version 16.1.0 Release 16)", Technical Report, 2020.

[45] J. Navarro-Ortiz *et al.*, "A Survey on 5G Usage Scenarios and Traffic Models", *IEEE Communications Surveys & Tutorials*, vol. 22, pp. 905–929, 2020 (https://doi.org/10.1109/COMST.2020.2971781).

[46] A. Iqbal, T. Khurshaid, A. Nauman, and S.-B. Rhee, "Energy-aware Ultra-reliable Low-latency Communication for Healthcare IoT in Beyond 5G and 6G Networks", *Sensors*, vol. 25, art. no. 3474, 2025 (https://doi.org/10.3390/s25113474).

Prashant Kumar Baheti, Research Scholar
Department of Computer Science and Engineering
https://orcid.org/0009-0003-0728-0009
E-mail: 2020phdevenprashant9185@poornima.edu.in
Poornima University, Jaipur, Rajasthan, India
https://www.poornima.edu.in

Ajay Khunteta, Ph.D., Professor
Department of Computer Science and Engineering
https://orcid.org/0000-0002-5335-9434
E-mail: ajay.khunteta@poornima.edu.in
Poornima University, Jaipur, Rajasthan, India
https://www.poornima.edu.in