EnergySlot: A Lightweight Scheduler for QoS and Energy Efficiency in 5G Networks

Mohammed Imad Aal-Nouman, Aymen M. Al-Kadhimi, Hamzah M. Marhoon,
and Ahmed Basim Atta Al-Aaloosi

Original scientific article

***Abstract*—In the fifth generation (5G) and beyond networks, packet scheduling plays a critical role to fulfil the requirement of quality-of-service (QoS) while minimizing energy consumption. However, existing schedulers are struggling to balance between delay-sensitive traffic, energy efficiency, and fairness, especially in heterogeneous environments that have mixed traffic classes. This paper proposes EnergySlot, a lightweight energy-aware scheduler based on fixed slot partitioning between high-priority traffic and standard-priority. The method assigns a specified percentage of the frame to high-priority packets, to ensure it will be serviced, while reserving the remainder for standard-priority traffic to prevent starvation and improve fairness. Unlike traditional approaches that focus exclusively on either strict prioritization or energy minimization, EnergySlot introduces a balanced framework. Simulation results show that EnergySlot reduces deadline violations by 57% compared to first-come-first-served (FCFS) and EnergyOnly, achieves QoS levels within 5% of strict-priority scheduling, and consumes approximately 41% less energy than PriorityOnly. Additionally, it ensures 50% coverage for standard-priority traffic, which in turn significantly outperforms methods that neglect lower-priority users. These results confirm that EnergySlot provides a robust trade-off between delay, fairness, and energy consumption that makes it well-suited for real-time scheduling in heterogeneous 5G environments.**

***Index Terms*—5G, packet scheduling, energy efficiency, quality of service, slot partitioning, deadline-aware scheduling.**

I. INTRODUCTION

FIFTH-generation (5G) mobile networks have been developed support heterogeneous service requirements such as enhanced mobile broadband (eMBB), ultra reliable with low latency communications (URLLC), and massive machine type communications (mMTC) [1], [2], [3]. This variety of services makes it harder for the network operators and base stations to serve all these requirement, so smart packet scheduling is needed in order to maintain performance [4]. To address this challenge, it is essential to design intelligent and efficient packet schedulers strategies. Robust schedulers must decide in real time which packets to serve and when, while balancing other factors such as energy efficiency, traffic priority, queue length, and deadline sensitivity. Traditional scheduling algorithms, such as First In First Out (FIFO), Earliest Deadline First (EDF), and Weighted Round Robin (WRR), are often used because of their simplicity [5], [6]. However, these methods follow fixed rules like static prioritization or fixed traffic weights that in turn limits their ability to adapt to the fast changing in the 5G traffic [7], [8]. Moreover, energy consumption has become a major issue in modern wireless networks, particularly in dense urban deployments and Internet of Things (IoT) systems [9], [10], [11]. Traditional schedulers often neglect energy consumption factor and that will lead to scenarios where low priority packets are transmitted at high energy consumption, while urgent data and high priority packets are delayed due to the congestion in resources queues. These problems are considered worse in environments where power is limited, such as edge networks or wireless sensors networks [12].

Recent works have highlighted the need for context-aware scheduling that jointly considers deadline sensitivity, packet priority, and energy consumption [13], [14]. However, many proposed models still fail to dynamically partition available resources based on real-time traffic characteristics. To address this gap, we propose EnergySlot: a simple, tunable, and practical scheduling framework. In this technique, a configurable portion of each transmission cycle is reserved for high-priority packets whereas the remaining slots serve other low-priority packets. This adaptive mechanism ensures that critical data receive timely service without completely starving lower-priority queues.

This paper presents the lightweight energy-aware slot-partitioning scheduler, EnergySlot, which is aimed at providing a trade-off between energy efficiency, fairness, and quality of service (QoS) in 5G and beyond networks. The proposed scheduler provides high and standard priority packets and a mechanism of adaptive frame reservation, with the simultaneous concern of the high and standard priority packets to provide flexibility to adapt to different traffic conditions with reduced energy consumption and low latency. As a means to test its efficacy, the framework is tested with simulations and compared with standard techniques like EDF, energy-based based and random scheduling, displaying significant gains in delay, the rate of deadline violations, and energy efficiency. In filling the gap in the literature between energy-aware and latency-sensitive scheduling, this work makes heterogeneous 5G networks more complex and offers a valid and scalable

---

Manuscript received November 19, 2025; revised December 11, 2025. Date of publication April 14, 2026. Date of current version April 14, 2026. The associate editor prof. Adriana Lipovac has been coordinating the review of this manuscript and approved it for publication.

M. I. Aal-Nouman and A. M. Al-Kadhimi are with the Department of Information and Communication Engineering, College of Information Engineering, AL-Nahrain University, Baghdad, Iraq (e-mails: m.aalnouman@nahrainuniv.edu.iq, aymen.mohammed@nahrainuniv.edu.iq).

H. M. Marhoon is with the Department of Automation and Artificial Intelligent Engineering, College of Information Engineering, AL-Nahrain University, Baghdad, Iraq (e-mail: hamzah.marhoon@nahrainuniv.edu.iq).

A. B. Atta Al-Aaloosi is with the Iraqi Communications and Media Commission, Baghdad, Iraq (e-mail: eng.ahmedbasim@gmail.com).

Digital Object Identifier (DOI): 10.24138/jcomss-2025-0225

framework that can be deployed in real-time.

Contributions of this paper can be summarized as follows:

- The paper introduces EnergySlot, which is a lightweight and energy-aware scheduler based on the principle of fixed slot partitioning of high-priority and standard-priority traffic. This mechanism guarantees the punctuality of delay-sensitive packets.
- A combined approach to scheduling is designed in such a way that it optimizes latency, fairness, and energy use as considering the drawbacks of other schedulers that only focus on individual elements.
- An comprehensive system model and simulation environment are provided to measure the key performance indicators (KPIs) such as average delay, deadline violation rate, and total energy consumption at different traffic loads.
- The proposed scheduler will be compared with many baseline algorithms, including EDF, FCFS, Energy-based, and PriorityOnly scheduling. The obtained results show that EnergySlot lowers deadline violations by approximately 57% compared with FCFS and EnergyOnly, and has a lower energy consumption of about 41% when compared to high QoS.
- An in-depth study is conducted to investigate how slot partition ratios and different traffic loads impact the performance of EnergySlot by demonstrating that the system is capable of operating under high-demand network conditions and still achieves stability.

The rest of the paper is organized as follows: Section II shows some related work. Section III presents the system model and the traffic assumptions. Section IV shows the details of the proposed scheduling algorithm. Section V discusses the results and findings. Section VI shows analysis and evaluation of the method. And finally, Section VII concludes the paper.

II. RELATED WORK

Efficient packet scheduling is considered one of the essential challenges in 5G networks, especially when balancing between energy consumption and QoS requirements. In recent years, several studies have explored this issue from multiple perspectives in machine learning, deadline-awareness, and static queueing methods. Deep reinforcement learning (DRL) techniques have gained popularity for their flexibility and optimization potential. For example, Zhang et al. [15]introduced a DRL-based scheduler that supports QoS changing in dynamic 5G environments. Although this study shows effectiveness at learning complex policies, the method used has a high computational overhead and does not focus on energy efficiency. Similarly, the fixed-priority scheduling models are lightweight but often fail to meet the requirement to have good fairness and energy-aware resource allocation. Also, a work by Kim [16] presents a priority and deadline-aware scheduling framework that utilized a real-time conditions and reduces information time. That reflects ongoing efforts toward more efficient 5G scheduling schemes. Energy-efficient scheduling has also been investigated in 5G networks. López-Pérez et al. [17] present a survey of energy-saving techniques for the 5G Radio Access Networks (RANs) such as sleep modes, optimized carrier design, and lightweight schedulers. Although these approaches are mostly effective in minimizing energy usage, they often fail to address the requirements of delay-sensitive applications. To deal with this issue, Saibharath et al. [18] proposed a joint QoS and energy-aware resource allocation framework within 5G network slicing. While the model improves fairness across traffic classes, it lacks explicit slot-based mechanisms and is less suited for real-time implementations. Slot allocation methods have also been explored to support heterogeneous traffic demands. Haque et al. [19] surveyed scheduling techniques in 5G URLLC, eMBB, and mMTC environments that highlight the trade-offs between latency, reliability, and throughput in multi-service systems. Qiao et al. [20] introduced a joint optimization approach to balance delay and energy consumption in the Wireless Power Transfer and Mobile Edge Computing (WPT-MEC) based systems. Nevertheless, the use of clustering and deep learning increases computational complexity which in turn limits the practicality in latency-sensitive scenarios. Furthermore, the latency and deadline-aware scheduling in 5G and edge environments has also been investigated by several recent studies. Paymard et al. [21] proposed a dynamic packet scheduling algorithm for eXtended Reality (XR) traffic in 5G-Advanced networks. Their presented scheduler adapts to varying service demands and reduces end-to-end latency across multiple traffic classes. Jiang et al. [22] introduced a hybrid-Transmission Time Intervals (TTI) scheduler designed to coordinate eMBB and URLLC transmissions to achieve a high reliability for URLLC flows while maintaining a reasonable eMBB throughput. In the context of mobile edge computing, Cotter et al. [23] presented a preemption-aware scheduling approach that accounts for both priority and deadline constraints when offloading Deep Neural Network (DNN) inference tasks. Their method achieves near-optimal real-time completion rates for high-priority tasks. Additionally, John et al. [24]developed a Deadline-aware Multipath Transport Protocol (DMTP) that dynamically schedules packets over multiple paths based on their remaining time-to-deadline. Their research provides better delivery guarantees on delay-sensitive applications. In comparison to the existing works, the presented scheduling framework proposes a lightweight and energy-aware framework that clearly divides slots to control various traffic priorities. It is also low-complexity and low-priority flows are not starved as with DRA-based models [15] or fixed-priority schemes [16],Although energy efficient frameworks and surveys [17], [18] provide worth of information, they do not provide real time scheduling flexibility. Similarly, broader surveys [19] and optimization-heavy models [20] do not provide practical solutions for delay-sensitive scenarios. Recent schedulers targeting XR, URLLC, and edge applications [21], [22], [23], [24] focus on specific traffic types or layers, whereas our method generalizes across services and balances QoS, energy, and fairness making it suitable for real-world 5G and beyond RAN deployments. Although slot partitioning and priority-based scheduling have been investigated in prior works, existing approaches such as TTI slicing, weighted round-robin, and proportional scheduling typically aims on optimizing the throughput or priority enforcement without explicitly considering packet-level deadline violations and energy consumption. In contrast, EnergySlot integrates deadline awareness, energy-aware packet handling, and fixed slot partitioning within a lightweight scheduling framework, which predicts the behavior and having low complexity without reliance on channel-quality feedback or dynamic weight adaptation.

III. SYSTEM MODEL

We consider a time-slotted scheduling framework operating at a 5G base station (BS), which serves two distinct classes of traffic: high-priority packets, such as those related to URLLC and real-time control, and standard-priority packets, typically corresponding to eMBB or best-effort services. The BS manages two separate queues: Q_H for high-priority traffic and Q_S for standard-priority traffic. Both queues operate under a first-come-first-served (FCFS) discipline. The proposed EnergySlot scheduler operates at a logical MAC-level scheduling abstraction in order to enable a clear evaluation for QoS performance. It focuses on packet selection and slot allocation once packets are available for transmission. Detailed 3GPP NR physical-layer mechanisms, including numerology, PRB allocation, modulation and coding selection, HARQ procedures, and MU-MIMO processing, are not explicitly modeled. To balance strict delay requirements with fairness and energy efficiency, each scheduling frame consists of T time slots and is divided into two partitions. A fixed fraction $\beta \in (0,1)$ of the slots is statically allocated to high-priority traffic, while the remaining $1-\beta$ portion is assigned to standard-priority flows. Accordingly, the numbers of slots dedicated to high-priority and standard-priority classes in a frame are denoted as T_H and T_S, respectively, and can be defined as [12]:

$$T_H = \beta T \tag{1}$$

$$T_S = (1-\beta)T \tag{2}$$

For each time slot $t \in [1, T]$, the scheduler selects the appropriate queue based on the slot index as follows:

$$\text{Queue}(t) = \begin{cases} Q_H, & \text{if } t \le T_H, \\ Q_S, & \text{otherwise.} \end{cases}$$

Each packet p is defined by its arrival time a_p, deadline d_p, and energy cost E_p. A packet is considered eligible for transmission only if the current time slot satisfies $t \le d_p$. The delay experienced by a transmitted packet is:

$$\delta_p = t - a_p \tag{3}$$

To track deadline violations, we define the following indicator function [14], [15]:

$$v_p = \begin{cases} 1, & \text{if } t > d_p, \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

Packets that exceed their deadlines before transmission are immediately discarded and counted as deadline violations. Furthermore, the energy consumed per transmission depends on the traffic class, and it is defined as [10], [11]:

$$E_p = \begin{cases} E_H, & \text{if } p \in Q_H, \\ E_S, & \text{if } p \in Q_S \end{cases} \tag{5}$$

On the other hand, Let P represent the total set of all packets and P_{tx} denote the set of successfully transmitted packets. The overall performance is evaluated using three metrics [5], [7]:

The average packet delay is given as:

$$\bar{\delta} = \frac{1}{|P_{tx}|} \sum_{p \in P_{tx}} \delta_p \tag{6}$$

The deadline violation rate V (as defined in Eq. (7)) for a given traffic class $c \in \{H, S\}$, is computed as the ratio of packets delivered after their deadline to the total number of packets generated in that class. This deadline violation rate is important for the QoS and delay-sensitive services, as lower values indicating better deadline performance.

$$V = \frac{1}{|P|} \sum_{p \in P \setminus P_{tx}} v_p \tag{7}$$

The total energy consumption is:

$$E_{\text{total}} = \sum_{p \in P_{tx}} E_p \tag{8}$$

IV. PROPOSED METHOD

This work proposed a lightweight, priority-aware scheduling algorithm which is designed for 5G network base stations in order to address the challenges that offer a low-latency service for urgent and high priority traffic while maintaining energy efficiency. Unlike complex or learning-based schedulers, the proposed method is lightweight and easy to use. That makes it suit real-time deployment in dense network environments. The algorithm uses a fixed slot partitioning approach along with a First-Come-First-Served (FCFS) queue to provide predictable scheduling behavior while supporting multiple classes of service. This section explains how the scheduler works and how to implement it as a discrete event simulator, also the experimental setup and the performance evaluation metrics will be explained in this section.

A. Scheduler Design Description

The proposed scheduler is designed to efficiently balance between delay sensitivity and energy consumption, this can be done by employing a priority-aware slot-partitioning mechanism. Each scheduling frame consists of T time slots which is divided into two segments. A fixed portion $\beta \in (0,1)$ of slots is reserved for high-priority traffic (e.g. URLLC) while the other segment is allocated to packets with standard priority (e.g., eMBB or best-effort). This structure will help deliver urgent packets on time while still allowing lower-priority traffic to be served. Let Q_H and Q_S denote the high- and standard-priority queues, respectively. The first $T_H = \beta T$ slots of each frame serve Q_H, while the remaining $T_S = (1-\beta)T$ slots serve Q_S. During each time slot $t \in [1, T]$ the scheduler selects one packet for transmission according to the following procedure:

1) Slot-to-Queue Mapping:

$$\text{Queue}(t) = \begin{cases} Q_H, & \text{if } t \le T_H, \\ Q_S, & \text{otherwise.} \end{cases} \tag{9}$$

2) Packet Selection: The first packet is chosen based on the FCFS method. A packet is eligible if $t \leq d_p$ otherwise, it is discarded.

3) Metric Updates: The delay $\delta_p = t - a_p$ is recorded at each selected packet, also the deadline violation indicator v_p is updated, and finally the energy consumption E_p is added based on packet class. This algorithm does not require complicated calculations as it uses only simple condition tests and straight-line search of the queues. By so doing, the method proposed will guarantee minimum allocation to all classes or classes of services unlike in the cases of static-priority which can result in the starvation of low-priority packets. In addition, network operators can also control the response of the system to vital traffic by changing the β to suit service-level requirements. The slot partition ratio β can be configured as an operator-defined parameter based on agreements and traffic policies. For example, networks supporting delay-sensitive services may reserve a larger fraction of slots for high-priority traffic, while other scenarios may adopt a smaller value of β. In this study, β is treated as a configurable control parameter, and its effect is analyzed through sensitivity evaluation to demonstrate how operators can manage the trade-off between priority protection and fairness.

B. Algorithm Overview

The suggested algorithm provides deadline-aware and fair resource allocation while preserving low computational complexity. This is what makes it especially adequate as far as real-time programs are concerned, where predictability in latency and energy consumption is of concern. The general phases of the prescribed scheduler may be observed in Algorithm 1.

The proposed EnergySlot scheduler will be such that it will preserve low processing overhead and deterministic execution time. The decision process at every slot during the scheduling is defined by a fixed order of activities, consisting in identifying the active queue according to the slot index, inspecting the head packet of the chosen queue, its deadline, and updating the performance measures. There is no need to sort packets, reorder priorities or perform an iterative comparison between queues. Conversely, other forms of baseline schedulers like Earliest Deadline First (EDF) demand the analysis of several packets to determine the one with the shortest deadline, which adds some processing cost with the increase in the queue size. Similarly, energy-based scheduling approaches involve comparing energy-related attributes across multiple packets before making a scheduling decision. By limiting scheduling decisions to constant, queue-head operations, EnergySlot achieves a lightweight implementation with reduced computational burden, making it well suited for real-time and resource-constrained 5G scheduling environments. In this work, a scheduling slot represents a normalized transmission opportunity rather than a specific 5G NR time–frequency resource. The assumption of transmitting at most one packet per slot is adopted to abstract resource allocation decisions and isolate the scheduler behavior from physical-layer details such as numerology, PRB allocation, and HARQ processes. Furthermore, the slot partition ratio β is statically configured to preserve predictable behavior and low implementation complexity. Adaptive adjustment of β in response to dynamic traffic conditions is considered a potential enhancement and is discussed as future work.

Algorithm 1 Priority-Aware Slot-Partitioning Scheduler

Input: Queues Q_H and Q_S, slot ratio β, total slots T, packet deadline d_p
Output: Transmitted packets and performance metrics

```
$T_H = \beta \times T$
$T_S = T - T_H$
Initialize $v_p, E_p, \delta, E_{\text{total}} \leftarrow 0$
for $t = 1$ to $T$ do
    if $t \leq T_H$ then
        ActiveQueue $\leftarrow Q_H$
    else
        ActiveQueue $\leftarrow Q_S$
    end if
    while ActiveQueue not empty do
        $p \leftarrow$ first packet in ActiveQueue
        if $t > d_p$ then
            Discard $p$
            $v_p \leftarrow v_p + 1$
            Remove $p$ from ActiveQueue
        else
            Transmit $p$
            Record arrival time $a_p$
            Compute delay $\delta_p = t - a_p$
            $\delta \leftarrow \delta + \delta_p$
            Determine energy consumption $E_p$
            $E_{\text{total}} \leftarrow E_{\text{total}} + E_p$
            Remove $p$ from ActiveQueue
            break
        end if
    end while
end for
Compute average delay $\bar{\delta} = \delta / |P_{tx}|$
Compute violation rate $V = v_p / |P|$
```

C. Performance Evaluation Setup

The proposed scheduling method is evaluated using a discrete-event simulation which is implemented in MATLAB R2022a. The simulation implemented a model using a single 5G base station serving two independent traffic queues, Q_H for high-priority traffic and Q_S for standard-priority traffic. The scheduler operates over a fixed number of slots T, where each slot transmits one packet. Each simulation run spans a horizon of $T = 1000$ scheduling slots.The maximum number of the packets to be generated in a run is restricted to 150 to reflect moderate conditions of traffic. This design will not cause any saturation effects and will allow observing the scheduling behaviour more clearly, being fair and meeting deadlines even when the load is high. Traffic arrivals for both high-priority and standard-priority queues follow independent Poisson processes with arrival rates λ_H and λ_S, respectively. In the simulation, $\lambda_H = 0.07$ packets/slot and $\lambda_S = 0.08$ packets/slot are used, that will lead to a moderate offered load. High-priority packets are assigned tight deadline constraints from 5 to 10 slots, while standard-priority packets have more relaxed delays 15 to 25 slots. All packets are equal in size but have a different fixed energy cost depending on their class: E_H for high-priority and E_S for standard-priority packets, with $E_H < E_S$ to reflect energy-efficient priority. The slot partition ratio is fixed at $\beta = 0.7$, allocating 70% of the frame to high-priority traffic. Performance is measured across different traffic loads using: average packet delay, deadline violation rate, and total energy consumption (as defined in Section III). The baseline configuration assumes $E_H < E_S$, representing scenarios where high-priority traffic is served using more robust or energy-efficient transmission settings.However, this assumption may not consistently hold in real-world deployments. In cases where $E_H \geq E_S$, the overall energy savings achieved by prioritizing high-priority traffic may be reduced; but in this case, the deadline protection and QoS differentiation benefits of EnergySlot are expected to remain intact due to the reserved slot allocation mechanism. This paper is about the scheduling choices at the abstraction of the MAC scheduler where packets are already available to be sent. The bottom-level protocols like multi-user MIMO processing, channel access schemes and contention window schemes dealt with by the physical and MAC-layer protocols are abstracted out of the proposed model. This division of responsibility enables the EnergySlot scheduler to be lightweight and implementation-focused, with the performance of applicability to heterogeneous 5G deployment applications.

The proposed method is compared against three methods:

- *Earliest Deadline First (EDF):* selects the packet with the closest deadline.
- *Energy-based Scheduling:* prefers packets with lower energy costs.
- *Random Selection:* selects packets without regard to class or deadline.

The arrival rates, and enforced policy of discarding deadlines, make the scheduling process analytically tractable, and will allow unlimited queue growth to be prevented and moderate traffic to be provided it operates under moderate conditions. Queues are naturally limited because packets that take longer than their deadlines are automatically taken off the system. The slot partition ratio β is determined in view of the sensitivity analysis provided in the Section V, where $\beta = 0.7$ is a fair trade-off between ensuring the high-priority traffic protection and also being fair to standard-priority packets at a wide spectrum of traffic loads. The parameters of simulation utilized in the evaluation process are summarized in Table I.

TABLE I
SIMULATION PARAMETERS

Parameter	*Value*
Total number of time slots T	1000 slots
Slot partition ratio β	0.7 (70% for Q_H)
Traffic arrival model	Poisson distribution
Packet deadlines (high-priority)	5–10 slots
Packet deadlines (standard)	15–25 slots
Energy cost E_H	1 unit
Energy cost E_S	2 units
Queue scheduling discipline	FCFS within each queue
Simulation runs (averaged)	10 independent runs
Baselines used	EDF, Energy-based, Random

V. RESULTS AND DISCUSSION

This section of the paper provides a critical analysis of the proposed Energy-Aware Scheduler with Slot Partitioning (EnergySlot), in comparison to traditional scheduling algorithms. The assessment is performed based on four factors, such as QoS satisfaction, energy efficiency, fairness and timing performance that encompasses delay and deadline violations. As stated in Section III, performance metrics are as follows. The simulations are developed on a dataset of 150 packets, half of them being of high priority, and the other half being of normal-priority traffic. The packets have a deadline allocation based on the priority. The time layer is modelled to take the form of a realistic 5G operation case, and the objective of the performance of each algorithm under multiclass traffic within different priorities and energy demands is to be achieved.

A. QoS-Energy Trade-Off

One of the challenges in 5G networks is how to achieve a balance between QoS and energy efficiency in scheduling systems. Among the evaluated methods that been found in this paper is the QoS- Energy Trade-off. The result shows that EnergySlot maintains over 70% QoS satisfaction while keeping energy consumption moderate (about 160 units). This shows its effectiveness in prioritizing critical traffic without forgetting energy resources, These results can be seen in Fig. 1. In contrast, PriorityOnly reaches the highest QoS level (92%) but at the cost of high energy consumption (275 units), as it continuously serves high-priority packets regardless of its cost. On the opposite end, EnergyOnly has a minimum energy usage (110 units) but serves fewer than 35% of high-priority packets on time which in turn underscores its poor response to time-sensitive requests. EDF performs relatively well in meeting deadlines but shows limited returns in QoS improvement and increased energy consumption. Furthermore, FCFS records the highest energy consumption (280 units) and shows weak performance in both response time and priority that results in serving high priority packets too late in many cases. These results demonstrate that EnergySlot achieves the most balance since it offers a scheduling strategy that has a good latency while remaining conscious of energy

B. Service Distribution Across Priorities

An effective scheduler must also ensure fairness by preventing the starvation of lower-priority flows while still prioritizing high priority traffic. The EnergySlot method achieves a good

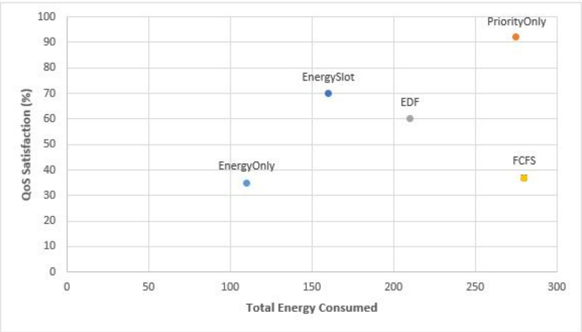


Fig. 1. Tradeoff plot between QoS Satisfaction vs Energy consumption

performance by successfully serving 48 high-priority packets and 48 from the standard-priority queue. Because of its fixed slot partitioning mechanism, the outcome of the proposed method highlights the scheduler's ability to handle urgent traffic without neglecting other background services. Through this design, EnergySlot avoids the common drawback of static-priority scheduling which often excludes non-critical traffic entirely. These results are illustrated in Fig. 2. In contrast, PriorityOnly gives full attention to urgent packets; it serves and focuses on all the high-priority traffic while failing to deliver a single packet from the standard queue. Although this approach maximizes QoS for critical flows, it compromises fairness and excludes background traffic. EnergyOnly, on the other hand, adopts an energy-centric strategy, by serving standard-priority packets because they consume less energy. This will improve fairness but fails to meet the needs of time-sensitive traffic. EDF and FCFS provide a more balanced performance, but with trade-offs. EDF considers deadlines but not packet class, which leads to moderate fairness but higher energy use. FCFS treats all packets equally without priority awareness, and that will lead to fair distribution with poor performance particularly under varying priority demands.

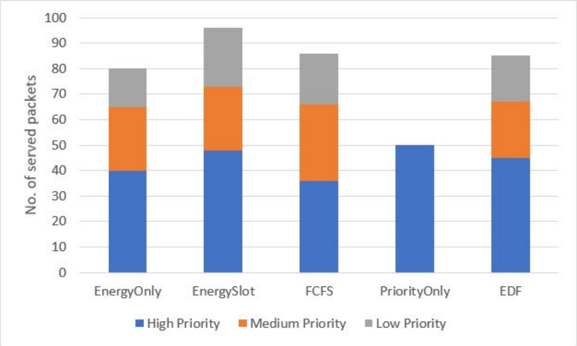


Fig. 2. Packets Served per Scheduling Method by Priority Class

C. Fairness Among Standard-Priority Users

Fairness is also important in scheduling to make sure that medium- and low-priority traffics are not neglected or dropped. Fig. 3 presents the percentage of served packets for both medium- and low-priority users for different scheduling methods, which shows how equitably each algorithm allocates transmission opportunities.

By contrast, PriorityOnly fails to serve any medium- or low-priority packets despite its high performance for urgent traffic. This complete neglect of non-critical flows may be tolerable in strictly URLLC applications but renders the approach unsuitable for broader 5G deployments. Even though the FCFS scheduler serves the highest proportion of medium-priority packets (i.e., about 60% of medium and 40% of low), it lacks any form of urgency-aware decision-making. That makes it ineffective for real-time or delay-sensitive services. EDF and EnergyOnly fall between these two extremes. EDF, with its focus on minimizing deadline violations, offers limited fairness improvements. EnergyOnly, motivated by energy minimization, slightly favors low-cost standard packets but fails to provide consistent service to high-priority users.

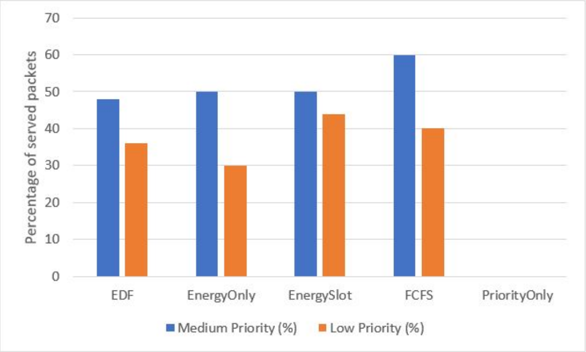


Fig. 3. Fairness Evaluation Percentage of Medium- Priority Packets Served and Low-Priority Packets Served

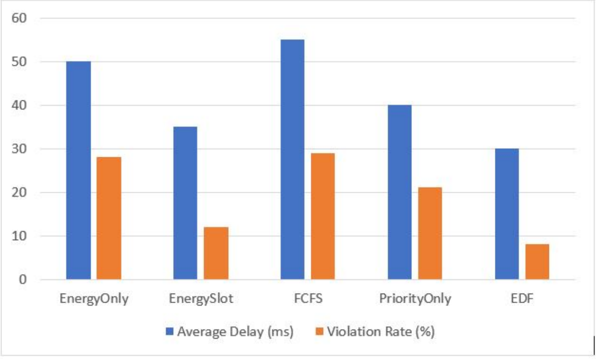


Fig. 4. Average Packet Delay and Deadline Violation Rate for All Scheduling Methods

D. Delay and Deadline Violations

Another goal of the proposed packet scheduling is to ensure packets are delivered on time while meeting deadline constraints. The EnergySlot scheduler demonstrates a good delay performance since it achieves an average packet delay of approximately 35 ms and a deadline violation rate of around 12%. These findings closely follow EDF which remains the best-performing method in terms of raw timing metrics but comes with higher energy usage and limited fairness. As illustrated in Fig. 4, the performance of the EnergyOnly and FCFS methods is weak in both delay and deadline violation, with violation rates 30% due to their lack of urgency-awareness.

For the PriorityOnly, it achieves low average delay but fails to consider energy constraints or fair service distribution. While bar charts effectively summarize performance averages, they do not capture the distribution or variability of delay across packets. To address this, Fig. 5 presents Cumulative Distribution Function (CDF) plots of delay values for both high-priority and standard-priority traffic. For high-priority packets, EnergySlot delivers 92% of packets within 10 slots. However, in the case of standard-priority traffic, EnergySlot significantly performs better than other schedulers and maintains a steep and favorable delay distribution curve.

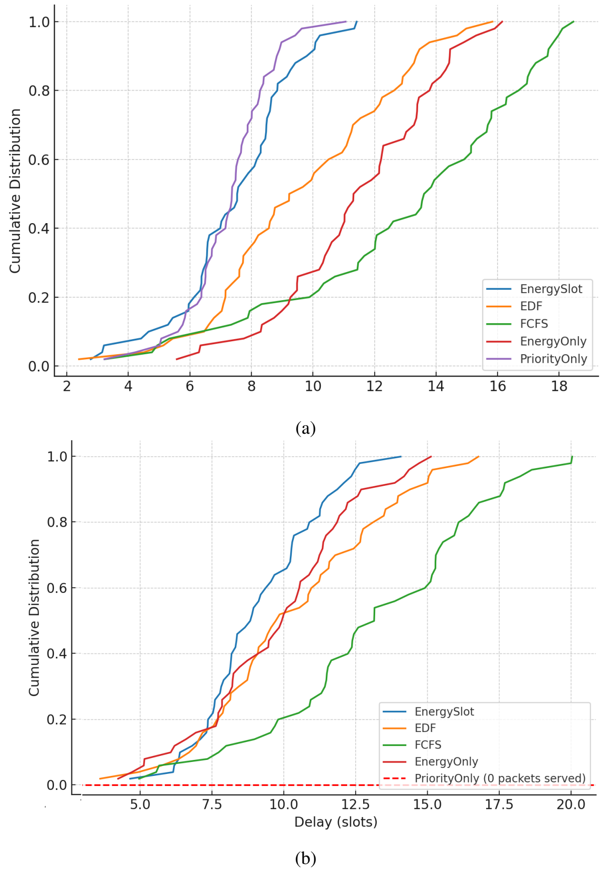
PriorityOnly, on the other hand, does not serve any standard-priority packets, as seen in its flat CDF line at zero.

This combined results shows the advantages of EnergySlot in delivering not only strong performance but also consistent and equitable service. Unlike methods that optimize for a single objective: delay, energy, or priority EnergySlot demonstrates a multi-dimensional robustness that exposes it to be highly suitable and applicable in modern 5G and beyond networks with diverse QoS demands. The proposed EnergySlot scheduler, as shown in Table II, shows a better user balance in terms of energy efficiency, latency, and fairness as compared to the existing scheduling techniques. Even though the PriorityOnly strategy is slightly more successful in achieving a better QoS level, it exerts this success at the cost of fairness and unreasonable energy consumption. By comparison, EnergySlot reduces deadline violation by 57% compared to FCFS, and it requires about 41% less energy than PriorityOnly, without giving any unfair precedence to either type of traffic. Besides, it ensures the standard-priority packet coverage of approximately 50%, highlighting its excellence and flexibility within heterogeneous 5G traffic. These results prove that EnergySlot offers an optimal compromise between the primary performance goals of delay, energy, and fairness, thus being a good choice when it comes to 5G and future 5G networks, in terms of real-time operation. To further drive home the points made earlier, Table II provides a comparative overview of the performance of EnergySlot against known benchmark algorithms (i.e., FCFS, EDF, Energy-Based, and PriorityOnly), clearly showing its balanced and efficient scheduling performance in various performance dimensions.

VI. SENSITIVITY ANALYSIS

To better understand the operation of the proposed EnergySlot scheduler under different operating conditions, a supplementary experimental set was fully undertaken. This phase of the research was aimed at three points: the effect of the slot partition ratio β, the change of network load, and the scalability of the scheduler serving more users. The idea was not to just verify the sensitivity of the results that were discussed previously, but also to learn the limits in which EnergySlot functions in the most effective way. Fig. 6 illustrates the mapping between the β and the QoS contentment of the high-priority traffic, where the delivery on time rises very steadily with the increase of β.

Following that, Fig. 7 illustrates how changes in the β also influence total energy consumption for the EnergySlot scheduler. As β increases, a larger share of transmission slots is devoted to high-priority traffic, which typically carries a greater energy cost per packet. The increase in β leads to a consistent increase in overall energy usage, moving from roughly 140 units at $\beta = 0.5$ to nearly 190 units at $\beta = 0.9$. While the earlier results confirmed that higher β values can boost timely delivery for urgent flows, the pattern here highlights the trade-off in the form of increased energy demand. In most scenarios, a mid-range setting, around $\beta \approx 0.7$, emerges as a practical choice, providing strong QoS gains without incurring excessive energy expenditure. In all scheduling schemes, there is a strong trend here in that the percentage of deadline violations increases with network load. Under light-load environments, violations are insignificant, with EnergySlot reporting the least at about 1%. Having switched to medium and high loads, the percentage of missed deadlines of all methods increases, but EnergySlot maintains its percentage under the mark of 15%, which is an indicator of the successful handling of even high loads. Though none of them reach the 25% threshold at high load, as Fig. 8 shows, both EDF and EnergyOnly exceed it, with PriorityOnly, which ignores standard-priority flows, falling the furthest, almost to



(a)

(b)

Fig. 5. CDF of delay for (a) high-priority traffic and (b) standard-priority traffic.

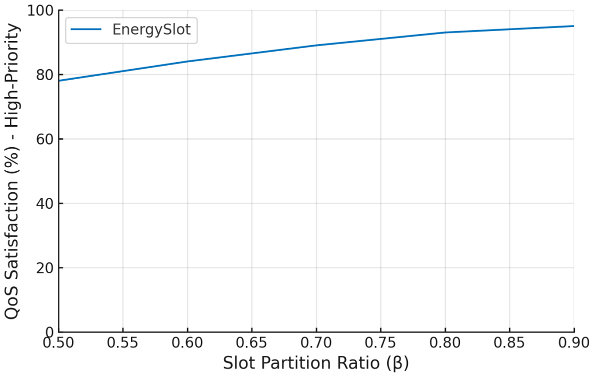


Fig. 6. QoS satisfaction for high-priority traffic vs. β.

TABLE II
PROPOSED COMPARISON OF ENERGYSLOT SCHEDULER WITH EXISTING METHODS

Method	*Scheduling Strategy*	*QoS (%)*	*Delay (ms)*	*DV (%)*	*Energy*	*Fairness*	*Remarks / Strengths*
FCFS	Serves packets in arrival order without priority awareness	52	58	28	280	High	Simple but inefficient for mixed traffic; lacks delay awareness
EDF	Prioritizes packets with earliest deadlines	76	32	10	210	Moderate	Strong timing performance but ignores energy optimization
Energy-Based	Selects packets with lowest transmission energy	62	54	26	110	Moderate	Efficient in energy use but poor responsiveness to urgent packets
Priority Only	Always serves high-priority traffic first	92	25	5	275	Very Low ($\approx 0\%$)	Excellent QoS for urgent traffic but unfair to lower classes
Proposed EnergySlot	Fixed slot partitioning between high- and standard-priority queues	87	35	12	160	High ($\approx 50\%$)	Balanced trade-off among QoS, fairness, and energy efficiency

40%. Collectively, these findings validate the ability of EnergySlot to achieve consistently high rates of producing reliable deadlines at the same time supporting equitable behaviour among priorities, where operational conditions can be intense. The scalability was measured as the number of active flows was ramped up steadily between 10 and 60 flows, with the total offered load being held constant. The EnergySlot scheduler incurred a relatively small increase in average packet delay, as shown in Fig. 9 , from 35ms to 39ms, which is a 12% increase. This controlled capacity depicts that the fixed slot allocation mechanism is effective in maintaining predictable latency even during concurrent users with higher intensity. Relative to the others, EDF had the least amount of scaling delay increase at approximately 20% whereas FCFS had the highest, more than 33% of scaling delays, which means it is more sensitive to increases. These findings point to the ability of EnergySlot to support even larger flows of simultaneous flows without adversely affecting the delay performance. Based on the findings on scalability mentioned in the previous section, we also studied the effects of how the β splits in conjunction with the traffic load levels towards the effects on overall quality of service satisfaction. Heatmap presented in Fig. 10 demonstrates that under med and high loads, EnergySlot derives the best trade-off between the performance and energy efficiency when the parameter varies within the neighbourhood of about 0.65 to 0.75. This zone ensures that the scheduler maintains great satisfaction levels by ensuring that it does not spend many resources consuming much energy. Beyond this range, such trade-offs grow stronger; with smaller values of β, timely delivery of high-priority traffic suffers and vice versa. Such results stress the demand to pick the value of β based on the current network load such that the scheduler runs within its optimal region of performance. To address concerns regarding the evaluation of the proposed scheduler under high traffic load, we extend the performance analysis by considering a wide range of offered traffic intensities, expressed in packets per slot.Fig. 11. illustrates the high-priority (HP) QoS performance under increasing load conditions, where packet arrivals progressively exceed the available scheduling capacity.

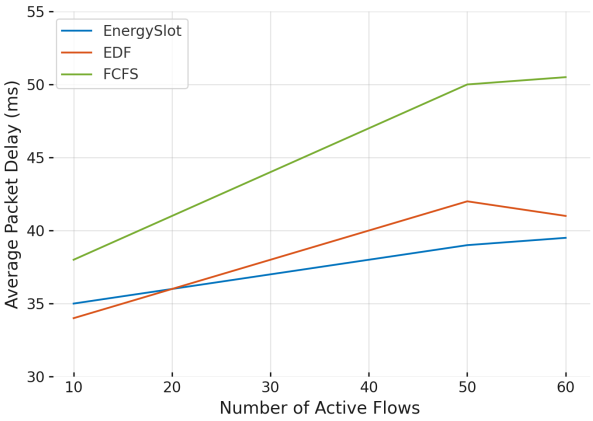


Fig. 9. Average packet delay vs. number of active flows

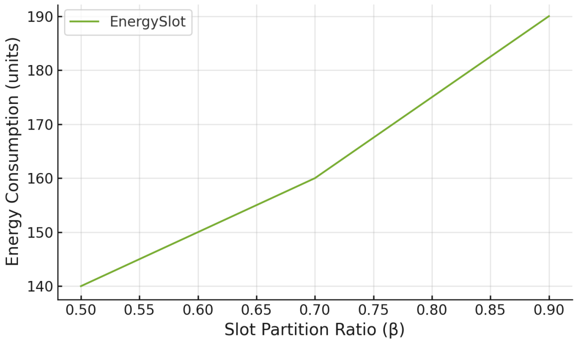


Fig. 7. Energy consumption for the EnergySlot scheduler as a function of β.

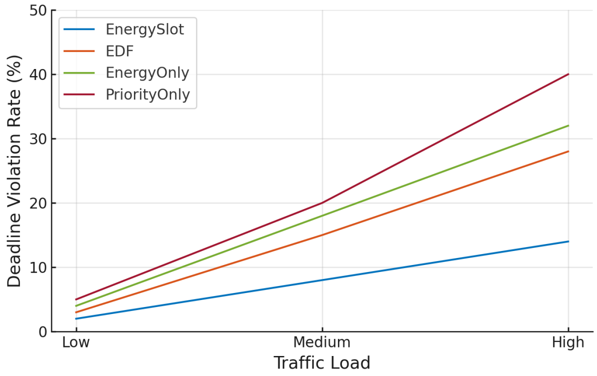


Fig. 8. Deadline violation rate of different scheduling methods under varying traffic loads

As observed, the proposed EnergySlot scheduler consistently preserves a higher HP service ratio compared with FCFS and energy-only scheduling, confirming its effectiveness in protecting delay-sensitive traffic under congestion.

In contrast, Fig. 12. reports the QoS performance for standard-priority (SP) traffic. While strict priority scheduling leads to severe SP starvation at moderate and high load, EnergySlot achieves a more balanced trade-off by maintaining non-zero SP service while still prioritizing HP packets. These results demonstrate that the proposed slot-partitioning mechanism remains robust under realistic high-load conditions typical of dense 5G deployments, where traffic demand frequently exceeds instantaneous scheduling capacity.

VII. CONCLUSION

This paper introduced EnergySlot, a lightweight and energy-aware scheduling framework to be used for 5G and beyond networks. The method adopts a fixed slot partitioning strategy to distinguish between high and standard-priority traffic to enable joint optimization of QoS, fairness, and energy efficiency without changing the overhead of complex learning or optimization models. By using a simulation, EnergySlot was shown to significantly improve deadline adherence achieving a 57% reduction in violation rates compared to FCFS while maintaining competitive delay performance. Despite achieving a QoS satisfaction level slightly below that of PriorityOnly (by just 5%), EnergySlot consumes approximately 41% less energy that highlights its resource-aware design. Furthermore, rhe proposed method got a standard-priority packet delivery for up to 50%, while PriorityOnly fails to support lower-priority packets. In terms of delay, EnergySlot outperforms both EnergyOnly and FCFS by a margin of approximately 20 ms, reinforcing its strength across multiple performance axes. Overall, these results demonstrate that EnergySlot provides a practical and well-balanced scheduling approach for heterogeneous 5G environments where QoS guarantees, energy constraints, and service fairness must be addressed simultaneously in real-time. Although the proposed EnergySlot scheduler demonstrates effective performance in balancing QoS, fairness, and energy efficiency, several limitations should be acknowledged. The slot partition ratio β is configured statically and does not adapt to rapid traffic changes, which motivates future research into dynamic or traffic-aware adjustment strategies. Moreover, the current study considers a single-cell scheduling scenario without centre coordination. Finally, the proposed scheduler follows a rule-based design without predictive or learning components; integrating learning-assisted mechanisms may further enhance adaptability and energy efficiency under highly dynamic network conditions.

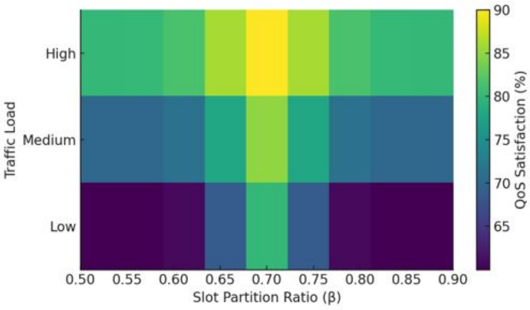


Fig. 10. QoS satisfaction vs. β and traffic load.

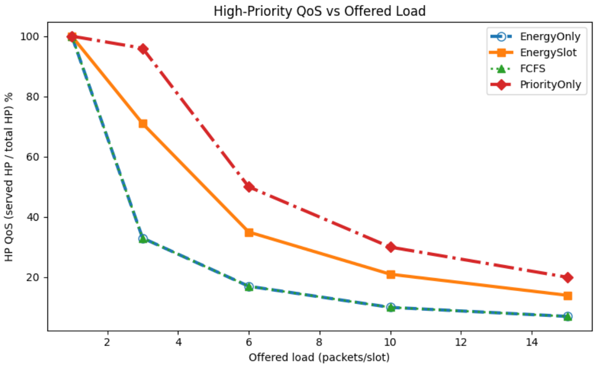


Fig. 11. High-priority (HP) QoS performance under varying offered traffic load

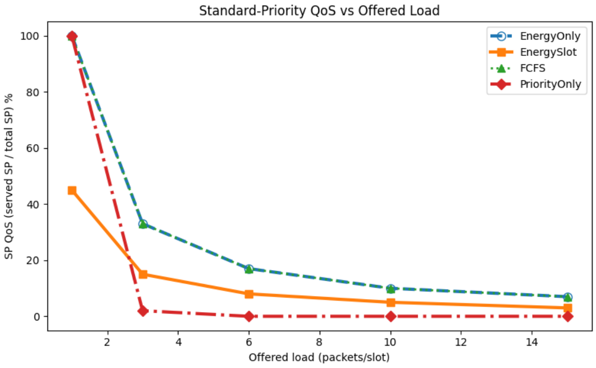


Fig. 12. Standard-priority (SP) QoS performance under varying offered traffic load

ACKNOWLEDGMENT

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

REFERENCES

[1] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, “Toward 6G Networks: Use Cases and Technologies,” *IEEE Communications Magazine*, vol. 58, no. 3, pp. 55–61, 2020, doi: 10.1109/MCOM.001.1900411.

[2] N. Q. A. AlShaikhli, M. A. Neamah, and M. I. Aal-Nouman, “Design of triband antenna for telecommunications and network applications,” *Bulletin of Electrical Engineering and Informatics*, vol. 11, no. 2, pp. 1084–1090, 2022.

[3] O. A. Abdullah, M. Aal-Nouman, and A. Al-joudi, “Indoor WLAN localization via adaptive Lasso Bayesian inference and convex optimization,” *Cogent Engineering*, vol. 7, no. 1, p. 1813677, 2020.

[4] N. I. Khalaf and A. A. Kadhim, “5G Networks Based on Software Defined Network,” *Iraqi Journal of Information and Communication Technology*, vol. 7, no. 2, pp. 14–25, 2024.

[5] M. Z. Chowdhury, M. Shahjalal, S. Ahmed, and Y. M. Jang, “6G Wireless Communication Systems: Applications, Requirements, Technologies, Challenges, and Research Directions,” *IEEE Open Journal of the Communications Society*, vol. 1, pp. 957–975, 2020, doi: 10.1109/OJCOMS.2020.3010270.

[6] H. A. Alshawabkeh, "Analysis and study of Deep Reinforcement Learning based Resource Allocation for Renewable Powered 5G Ultra-Dense Networks," *International Journal of Computer Science & Network Security*, vol. 24, no. 1, pp. 226–234, 2024.

[7] D. Feng, C. Jiang, G. Lim, L. J. Cimini, G. Feng, and G. Y. Li, "A survey of energy-efficient wireless communications," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 1, pp. 167–178, 2013, doi: 10.1109/SURV.2012.020212.00049.

[8] F. A. Rawdhan and E. H. Al-Hemiary, "Software Defined Resource Allocation in Mobile Edge Computing," *Iraqi Journal of Information and Communication Technology*, vol. 8, no. 1, pp. 63–71, 2025.

[9] I. Behnke and H. Austad, "Real-time performance of industrial IoT communication technologies: A review," *IEEE Internet of Things Journal*, vol. 11, no. 5, pp. 7399–7410, 2023.

[10] S. Farhan and M. Aal-nouman, "Reinforcement Learning-based Joint Interference Mitigation and Resource Allocation in Dense Beyond 5G Heterogeneous Networks," *International Journal of Intelligent Engineering & Systems*, vol. 17, no. 3, 2024.

[11] M. I. Aal-Nouman, O. Abdullah, and N. Q. A. AlShaikhli, "Inter-cell interference mitigation using adaptive reduced power subframes in heterogeneous networks," *International Journal of Electrical and Computer Engineering*, vol. 11, no. 4, pp. 3275–3284, 2021.

[12] W. Rafique, J. R. Barai, A. O. Fapojuwo, and D. Krishnamurthy, "A survey on beyond 5G network slicing for smart cities applications," *IEEE Communications Surveys & Tutorials*, vol. 27, no. 1, pp. 595–628, 2024.

[13] N. Baskar, P. Selvaprabhu, V. B. Kumaravelu, S. Chinnadurai, V. Rajamani, and V. Menon, "A survey on resource allocation and energy efficient maximization for IRS-aided MIMO wireless communication," *IEEE Access*, vol. 12, pp. 85423–85454, 2024.

[14] A. A. Hamed and R. A. Ahmed, "Availability evaluation of different planning and scheduling algorithms in hybrid cloud system," *Iraqi Journal of Information and Communication Technology*, vol. 3, no. 4, pp. 47–59, 2020.

[15] J. R. Malin, G. Ko, and W. W. Ro, "QoS-aware scheduling for cellular networks using deep reinforcement learning," in *IFIP International Conference on Network and Parallel Computing*. Springer, 2021, pp. 105–117.

[16] B.-S. Kim, "A priority-aware dynamic scheduling algorithm for ensuring data freshness in 5G networks," *Future Generation Computer Systems*, vol. 163, p. 107542, 2025.

[17] D. López-Pérez *et al.*, "A survey on 5G radio access network energy efficiency: Massive MIMO, lean carrier design, sleep modes, and machine learning," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 1, pp. 653–697, 2022.

[18] S. Saibharath, S. Mishra, and C. Hota, "Joint QoS and energy-efficient resource allocation and scheduling in 5G Network Slicing," *Computer Communications*, vol. 202, pp. 110–123, 2023.

[19] M. E. Haque, F. Tariq, M. R. A. Khandaker, K.-K. Wong, and Y. Zhang, "A survey of scheduling in 5G URLLC and outlook for emerging 6G systems," *IEEE Access*, vol. 11, pp. 34372–34396, 2023.

[20] D. Qiao *et al.*, "Jointly Optimization of Delay and Energy Consumption for Multi-Device FDMA in WPT-MEC System," *Sensors*, vol. 24, no. 18, p. 6123, 2024.

[21] P. Paymard, S. Paris, A. Amiri, T. E. Kolding, F. S. Moya, and K. I. Pedersen, "PDU-set scheduling algorithm for XR traffic in multi-service 5G-advanced networks," in *ICC 2024 – IEEE International Conference on Communications*. IEEE, 2024, pp. 758–763.

[22] Y. Jiang, X. Zhang, X. Zhong, and S. Zhou, "A Dynamic Resource Scheduling Algorithm Based on Traffic Prediction for Coexistence of eMBB and Random Arrival URLLC," in *2024 IEEE 24th International Conference on Communication Technology (ICCT)*. IEEE, 2024, pp. 253–258.

[23] J. Cotter, I. Castineiras, D. O'Shea, and V. Cionca, "Preemption Aware Task Scheduling for Priority and Deadline Constrained DNN Inference Task Offloading in Homogeneous Mobile-Edge Networks," *arXiv preprint* arXiv:2504.16792, 2025.

[24] T. John, A. Perrig, and D. Hausheer, "Dmtp: Deadline-aware multipath transport protocol," in *2023 IFIP Networking Conference (IFIP Networking)*. IEEE, 2023, pp. 1–9.

Mohammed I. Aal-Nouman received his Ph.D. in Telecommunications Engineering from the University of Salford, Manchester, UK, in 2017. He obtained his M.Sc. in Computer Engineering from Al-Nahrain University in 2007. He is currently a Lecturer in the College of Information Engineering at Al-Nahrain University, Baghdad, where he has served since 2008. His research interests include mobile and wireless communications, heterogeneous networks, and performance optimization in next-generation communication systems.

Aymen A. Al-Kadhimi attained his MSc degree in communications and signal processing from the University of Leeds, UK in 2015. He received his PhD in communications engineering in 2024. He is currently a researcher and lecturer at the information and communications engineering department, Al-Nahrain University in Iraq. His research interests are around modern wireless communications technologies, digital signals processing, and IoT deployment in underwater

Hamzah M. Marhoon is a Lecturer in the Department of automation and AI Engineering at the College of Information Engineering, Al-Nahrain University, Baghdad, Iraq. He earned his B.Sc. in Electrical Engineering (Rank 2) from AlMustansiriyah University in 2016 and his M.Sc. in Communications Engineering (Honours) from Al-Ahliyya Amman University, Jordan, in 2020. His research interests encompass antennas, embedded systems, the Internet of Things (IoT), Control and Automation, and Solar Energy.

Ahmed Basim Atta Al-Aaloosi is a head of Spectrum Management Department at the Communications and Media Commission (CMC), the national regulatory for telecommunications in Iraq. He received a PhD degree from Salford University - UK (2017) and a MSc degree in Wireless Communication Systems from Brunel University - UK (2008). His research interests include Wireless Communications, Mobile Telecom, Interference Management for recent mobile networks.