Simulation Modelling Practice and Theory 125 (2023) 102745

Contents lists available at ScienceDirect

Simulation Modelling Practice and Theory

journal homepage: www.elsevier.com/locate/simpat

System analysis of QoS schedulers for XR traffic in 5G NR

Katerina Koutlia [a,*], Biljana Bojovic [a], Sandra Lagén [a], Xiaodi Zhang [b], Ping Wang [b], Jimin Liu [b]

[a] Centre Tecnològic de Telecomunicacions de Catalunya (CTTC/CERCA), Avinguda Carl Friedrich Gauss, 7, 08860 Castelldefels, Barcelona, Spain
[b] Reality Labs, Meta, USA

ARTICLE INFO

Keywords:
ns-3
5G-LENA
QoS schedulers
XR traffic
System-level simulation
5G NR

ABSTRACT

Fifth Generation New Radio (5G NR) presents a new framework for Quality-of-Service (QoS) management that especially fits cases for eXtended Reality (XR) traffic applications thanks to the new presented concept of QoS flows that can handle various service data flows. In that respect, in this paper we exploit the QoS flow model and we present a QoS MAC scheduler that can classify XR and mixed traffic based on the assigned QoS indicators to each traffic. We perform an end-to-end system level evaluation under two scenarios using the ns-3 5G-LENA simulator. First, we validate the correct operation of the QoS MAC scheduler in a single-cell scenario under non-saturation and saturation conditions. Then, we consider a realistic multi-cell scenario defined by the ITU-R and used for calibration of the 5G-LENA simulator, in order to evaluate the proposed solution and study its performance under realistic network conditions. Results demonstrate the correct functionality of the proposed QoS MAC scheduler and prove its ability to classify XR and other types of traffic based on the desired QoS indicators.

1. Introduction

With the latest technology advancements and innovations, virtual experience using immersive solutions, such as eXtended Reality (XR), is becoming tangible. XR is envisioned to change the way entertainment, education, healthcare and manufacturing, among others, is known up to now [1,2]. Many applications can benefit from the potentiality of XR, constituting it as the next big computing platform, gaining a lot of momentum in the industry and consumer market. Examples include virtual tours, gaming, remote training and remote control in cases of risky and inaccessible environments [3,4]. All these applications have very stringent requirements in terms of bit rate and latency in order to offer a high quality immersive user experience. In fact, most of XR use cases are considered as time critical [1]. With the advances however, offered by the Fifth Generation New Radio (5G NR), such as time-constraint capabilities, improved throughputs and high reliability communications, the way towards the virtual world is widely opened [5].

XR and Cloud Gaming (CG) are considered to be some of the most important immersive technologies in the industry, with XR being widely used as a general term that includes various types of applications of virtual reality, as well as applications that combine real and virtual environments [6,7]. These are Virtual Reality (VR), Mixed Reality (MR) and Augmented Reality (AR). All these applications are characterized by different traffic requirements. As such, in the scheduling process it is crucial to properly manage each traffic type according to its requirements in order to achieve the desired throughput, latency and user experience. The differentiation can be achieved thanks to the introduction of a newly presented framework in the 5G technology, known as the 5G

* Corresponding author.
E-mail addresses: katerina.koutlia@cttc.es (K. Koutlia), biljana.bojovic@cttc.es (B. Bojovic), sandra.lagen@cttc.es (S. Lagén), xiaodizhang@meta.com (X. Zhang), pingwang@meta.com (P. Wang), jiminliu@meta.com (J. Liu).

https://doi.org/10.1016/j.simpat.2023.102745
Received 13 December 2022; Received in revised form 31 January 2023; Accepted 2 March 2023
Available online 15 March 2023

Quality of Service (QoS) model, that manages and controls the QoS requirements based on the concept of *QoS flows* [8]. In fact, this can be particularly beneficial for XR and CG traffic management, where the applications generate multiple service data flows with varying QoS requirements [5].

Although 5G provides the architecture to support XR traffics, MAC scheduler algorithms are not defined by the standard because they are implementation-specific functions. Therefore, the research question we intend to address in this paper is the following: Given the 5G QoS architecture, what QoS-based scheduling algorithms and metrics can be applied at the MAC layer to schedule a variety of simultaneous XR traffic users and to provide and guarantee the different XR QoS requirements? Therefore, this paper studies how to handle efficiently XR traffic applications that are composed of multiple sub-flows with different QoS requirements and provides a system level evaluation under mixed traffic scenarios.

1.1. Related works

QoS schedulers have been extensively investigated for 5G NR usage scenarios. For example, the authors in [9] present a QoS-aware and channel-aware Radio Resource Management (RRM) framework in a multi-numerology 5G NR system that manages Best Effort (BE) and Guaranteed Bit Rate (GBR) traffic types, while in [10], a knowledge-assisted deep reinforcement learning algorithm for time-sensitive traffic in 5G networks is developed. A QoS prediction scheme, known as PreQoS, for vehicular scenarios is presented in [11], exploiting machine learning methods. A recent survey that reviews a variety of scheduling algorithms, along with a study of metrics, performance goals, objectives and application suitability for 5G use cases, can be found in [12].

However, work on 5G NR QoS management for XR traffic is currently ongoing, therefore related work is quite scarce. Additionally, works that consider mixed traffic type (AR/VR/CG + other types of traffic) and multi-flow traffic, limit even more the state-of-the-art list. An initial work for LTE networks is presented in [13]. The paper presents a new scheduling algorithm for real time traffic based on the packet and head-of-line (HOL) delays. The evaluation is carried out for H.264 Video Stream, VoIP and BE traffics using the LTE-Sim Simulator [14]. A scheduling algorithm, known as demand based proportional fairness (D-PF), that classifies VR and UHD UEs is proposed in [15] for 5G eMBB scenarios. The authors in [16] present a scheduling framework for mixed traffic scenarios, known as 5MART, where reinforcement learning and neural networks are applied in order to take the most appropriate scheduling decisions. In [17], a resource allocation scheme based on reinforcement learning is proposed for heterogeneous traffic that manages QoS provisioning. Moreover, although one can argue that work such as the above presented can be applied in XR and mixed traffic scenarios (i.e., XR + other types of traffic), these algorithms are not XR specific and cannot handle all the peculiarities of the XR traffic, such as multiple service data flows with different QoS requirements.

1.2. Contribution

Based on the above, in this paper we present and evaluate a QoS scheduler model that takes into account the QoS flow requirements of the various service data flows under XR and mixed traffic. The evaluation is carried out through End-to-End (E2E) system-level simulations using the open-source 5G-LENA simulator [18], recently calibrated in 3GPP reference scenarios [19]. We perform the evaluation in two steps. First we consider a single-cell scenario of 1 gNB and several UEs with mixed traffic (AR/VR/CG/VoIP), in order to validate the correct functionality of the scheduler. Then, we evaluate the new QoS MAC scheduler in a realistic scenario of multiple gNBs and UEs, in order to study the performance of the proposed solution under real network conditions. Let us highlight that the multi-cell scenario has been used for the calibration of the 5G-LENA simulator and is defined by ITU-R. We compare the new QoS MAC scheduler against traditional schedulers, such as the Round Robin (RR) and Proportional Fair (PF) algorithms.

1.3. Paper organization

The rest of the paper is organized as follows. Section 2 presents the main characteristics of the 5G-LENA simulator used for the implementation and evaluation of the proposed QoS MAC scheduler. Moreover, we give a discussion related to the calibration of the simulator and we detail the current MAC Layer. Section 3, includes an overview of the 5G NR QoS model, the description of the theory behind the proposed QoS MAC scheduler, the implementation details and a presentation of the scenarios used for the performance evaluation. The simulation results and the QoS MAC scheduler validation and evaluation are given in Section 4. Finally, Section 5 includes the most important conclusions and the envisioned future work.

2. 5G-LENA background

For the analysis and evaluation of the proposed QoS MAC schedulers for XR and mixed traffic scenarios, we have adopted the 5G-LENA module of the ns-3 [20] discrete-event network simulator. 5G-LENA is an open-source NR system-level simulator presented in [21]. It implements a high-fidelity full protocol stack and allows cross-layer evaluations and the study of E2E performances [19]. The main characteristics of the simulator are given in the next sections.

2.1. 5G-LENA introduction

The 5G-LENA simulator has been designed to support bandwidths from 400 MHz to 100 GHz, using the 3GPP spatial channel model TR 38.901 [22]. Its current architecture is based on Non-standalone mode, i.e. 5G RAN and 4G EPC. Moreover, it supports multiple numerologies leading in flexible and automatic configuration of the NR frame structure. Other key features of the simulator include: Frequency-Division Multiple Access (OFDMA)-based access with variable Transmission Time Interval (TTI) and Time-Division Multiple Access (TDMA)-based access, flexible MAC schedulers that simultaneously consider time- and frequency-domain resources (resource blocks and symbols) both for TDMA and OFDMA-based access schemes with variable TTI, Frequency Division Duplexing (FDD) and Time Division Duplexing (TDD) with flexible and configurable TDD patterns, Uplink grant-based access scheme with scheduling request and 3GPP-compliant buffer status reporting, NR-compliant processing timings (K1, K2), Bandwidth Part (BWP) managers and the architecture to support operation through multiple BWPs and multiple Component Carriers (CCs), new NR-compliant physical layer abstraction, considering LDPC codes for data channels, and MCS tables 1 and 2 (including up to 256-QAM), compatibility with 3GPP uniform planar arrays with directional and isotropic radiation, beamforming support with various analog beamforming methods, ideal and realistic beamforming with SRS-based channel estimates, and Dual-polarized MIMO support with rank adaptation algorithms.

2.2. 5G-LENA calibration

One of the main reasons for the adoption of the 5G-LENA simulator has been also the fact that it has been recently calibrated in 3GPP outdoor reference scenarios [19], while in the past, calibration of the simulator has been performed in indoor hotspot scenarios [21]. Results confirm the validity of the simulator's performance and the resemblance to that of 3GPP industrial simulators is proven. The outdoor scenarios that have been used for the calibration are defined in Report ITU-R IMT-2020 [23] and are the Rural-eMBB and Dense-eMBB. As such, for the evaluation of the QoS MAC scheduler for XR and mixed traffic we have opted to study a realistic multi-cell scenario using the Rural-eMBB and a single-cell scenario used for the validation of the correct functionality of the scheduler, following the Dense-eMBB configuration.

2.3. MAC layer

The 5G-LENA MAC layer was designed to support variable TTI, therefore the number of allocated resources (i.e., OFDM symbols and physical resource blocks (PRBs)) to a user is variable. It interacts directly with the PHY layer through a set of SAP APIs and indirectly with the RLC layer [24]. RLC sends to the MAC the Buffer Status Report (BSR) messages, based on which the scheduler will take the scheduling decisions.

5G-LENA supports both TDMA and OFDMA-based access schemes with variable TTI, while the scheduler API implementation follows the femto-forum specification [25]. In the current version, three MAC scheduling policies are supported, i.e., Round Robin (RR), Proportional Fair (PF) and Maximum Rate (MR). At the architectural level, up to date, there was a one-to-one mapping between Service Data flows (SDFs) and Data Radio Bearers (DRBs).

3. 5G-LENA towards QoS assessment

3.1. XR traffics overview

3GPP Release-17 includes a recent study on XR evaluations for 5G NR, presented in [7], where it defines the downlink and uplink models for AR, VR and CG applications, along with details for the deployment scenarios and the evaluation methodology, such as the simulation settings and the key performance indicators (KPIs). The study targets the identification of the limitations of current 5G NR networks towards the support of XR and CG applications. To perform the study, the following models are defined for the AR, VR and CG applications:

- AR Traffic model includes single-stream (video) model for the downlink and a multi-stream model for the uplink, including various sub-variants with one (video), two (pose/control + scene/video/data/audio) or three streams. For the case of AR uplink with three streams, two sub-models are considered; Model 3A uses aggregated streams for pose/control, scene/video and audio/data, while Model 3B considers pose/control, video I-stream and video P-stream.
- VR and CG traffic models include single stream (video) and multi-stream models (two streams, video + audio/data) for VR/CG downlink, and single stream for VR/CG uplink (using pose/control traffic). VR and CG models differ in the statistical characterization of the traffics.

3.2. 5G NR QoS overview

The 5G NR QoS model considers a two-level mapping of IP-packets of SDFs to suitable QoS flows performed by 5G Core (5GC) and QoS flows to DRBs performed by RAN [26]. This way, the 5GC and RAN functionality is decoupled [8]. In addition, with this type of mapping it is possible for the 5GC to perform Mx1 mapping, meaning that multiple SDFs can be mapped to one QoS flow. Similarly, but independently from the 5GC, the RAN can perform Nx1 mapping of QoS flows to DRBs, meaning that multiple QoS flows can be mapped to one DRB. For this, a new layer is introduced in the RAN known as Service Data Adaptation Protocol (SDAP), targeting the management of QoS flows [26].

With respect to the classification of the various traffic flows, the 5G NR QoS model supports both guaranteed bit rate (GBR) and non-guaranteed bit rate (non-GBR) QoS flows, as well as delay-critical GBR QoS flows. To differentiate among the various QoS flows, a QoS Flow ID (QFI) is used in an encapsulation header [26]. In addition, each flow is characterized by a QoS Profile that includes a 5G QoS Identifier (5QI), an Allocation and Retention Priority (ARP), a Reflective QoS Attribute for the case of non-GBR flows, or a Guaranteed Flow Bit Rate (GFBR), Maximum Flow Bit Rate (MFBR), Maximum Packet Loss Rate, Delay Critical Resource Type and Notification Control for the case of GBR flows. Notice that 5QI includes information with respect to the management of the flow, such as the resource type (GBR/non-GBR/Delay-Critical GBR), the Priority Level, the Packet Delay Budget (PDB), Packet Error Rate (PER), Averaging Window and Maximum Data Burst Volume [26]. 3GPP has specified various 5QIs in [8] to provide a standardized framework for services that have common characteristics and requirements and can therefore benefit from optimizing the signaling under standardized QoS characteristics. Some of these 5QIs can be used directly for the case of XR traffic when low latencies are not required, such as XR streaming applications [27]. However, the general architecture of XR will require low latencies, therefore, 3GPP is standardizing new 5QIs that are XR specific, such as 5QI 80 and 5QI 87.

3.3. New MAC scheduler

In order to perform the scheduling process based on the different traffic requirements the QoS MAC scheduler needs to select the users to be scheduled and assign the available resources, i.e. the available symbols and PRBs to them based on the QoS flow profile. As such, the scheduler should take into account the characteristics of each flow (i.e. the 5QI information, such as resource type, priority level, PDB, etc.) in order to take the most appropriate scheduling decisions so as to determine the users with highest priority and consequently allocate the necessary resources to them.

The new design of the proposed QoS MAC scheduler is based on the selection of scheduling weights. Without losing generalization to other types of traffic and in order to meet the strict requirements imposed by the inherent time-critical nature of the XR applications, such as the low latency, we propose the calculation of the scheduling weight (w), that denotes the priority for a DRB to be scheduled, to be as follows:

$$w = \begin{cases} (100 - P)\frac{r^{\gamma}}{R(\tau)} + F & \text{for non-GBR and GBR} \\ (100 - P)\frac{r^{\gamma}}{R(\tau)} D + F & \text{for delay-critical GBR} \end{cases} \tag{1}$$

where P is the default Priority Level of the QoS flow carried in the DRB, r is the instantaneous achievable data rate calculated by the spectrum efficiency and the channel bandwidth, $R(\tau)$ is the past average data rate updated within updated window size τ, $F = 100$ when the DRB has retransmission data and $F = 10$ otherwise, and γ is a configurable parameter. Moreover, we include the newly introduced delay budget factor D, that is the delay-aware weight related to HOL packet delay and the PDB, and calculated as D = PDB/(PDB-HOL). It is worth highlighting that D is an original and important factor that allows the proposed scheduler to meet the low latency requirements of various traffic types with special focus on the XR traffic requirements. Notice that, although there is a vast amount of works in the literature that adopt various scheduling metrics based on the P, PDB, HOL and PF metric, e.g. [13], to the best of our knowledge there is no work that considers all these metrics in conjunction. This combination results in promoting flows with tighter latency budget and consequently it avoids packet discarding in these cases, while at the same time allows treatment of the flows according to their priority level and maintains the PF metric to ensure higher achievable throughputs and user fairness.

Note that in Eq. (1) a lower Priority Level value (P) indicates higher priority for scheduling. The past average data rate is calculated as:

$$R(\tau) = (1 - \alpha)R(\tau - 1) + \alpha A(\tau) \tag{2}$$

where $A(\tau)$ is the current data rate over updated window size τ computed as the ratio of all successful delivered bits (including those bits still in retransmission) in the past updated window size, and α balances between the current data rate ($A(\tau)$) and the past average data rate in the previous window ($R(\tau - 1)$). When $\gamma = 1$, and assuming same priority for all users and ignoring D, the scheduler corresponds to a typical Proportional Fair scheduler.

Moreover, let us point out that for the case of XR traffic the mapping process described in Section 3.2 is particularly suitable when the application generates multiple flows. In such case, a user may have one of more active DRB(s) in a given moment. When there is only one active DRB, the QoS MAC scheduler considers the scheduling weight (scheduling priority) of that DRB. When the user has multiple active DRBs, there are various design options for the calculation of the scheduling weight. For instance, the sum of all weights for each DRB can be considered or in a simpler form, the Priority Level of the highest priority flow among its active DRBs can be taken into account.

The users are then ordered descendingly in each TTI according to their scheduling weight. This means that users with higher w are scheduled first.

For the non-GBR and GBR QoS flows in the DRBs per user, the minimum number of bits allowed is computed by the token bucket algorithm according to the minimum flow bit rate. In case of delay-critical GBR QoS flow, such value is capped by the maximum data burst volume. For each candidate user with GBR QoS flow and/or delay-critical GBR flow, DRBs are allocated to accommodate the minimum number of bits one-by-one in order, until all PRBs are used or all GBR QoS flows are served. Then, for each candidate user with non-GBR QoS flows, the number of PRBs is computed according to the minimum number of bits out of their RLC buffers, until all PRBs are used or the RLC buffers are empty.

3.4. 5G-LENA implementation

For the implementation of the proposed QoS MAC scheduler we have extended the 5G-LENA simulator with new classes for both TDMA and OFDMA access modes. In particular, we have implement a TDMA QoS scheduler in the `NrMacSchedulerTdmaQos` class, and an OFDMA QoS scheduler in the class `NrMacSchedulerOfdmaQos`. These classes are used to set the scheduler and access mode types when desired by the user. Moreover, they update the Downlink (DL) and Uplink (UL) metrics for each UE and the potential throughput based on the available resources. This information will be used by the newly implemented class `NrMacSchedulerUeInfoQos` to perform the UE sorting (weights computation), based on the priority level of the assigned 5QIs (lowest 5QI has higher priority). Notice that in the current implementation the scheduler does not make distinction among the resource type. The simulator has been also extended to support Mx1 mapping of SDFs to DRB (the architecture details are out of scope of this paper). In such case, there is a single DRB, and the scheduling weight considers its assigned Priority Level. In the 1×1 mapping architecture, there may be multiple active DRBs per user, and the scheduling weight considers the Priority Level of the highest priority DRB for each user.

To carry out the evaluation of the QoS MAC scheduler, we have also implemented in the 5G-LENA simulator the XR traffic models (i.e., AR, VR and CG), as described in Section 3.1, including pose/control, audio/data and generic scene/video models. Although we consider that the implementation details of the XR traffic models are out of scope of this paper, it is worth noting that said models can be parametrized to configure the various XR applications under single and multi stream setups. Also, let us notice that we have extended the simulator to support both the XR specific 5QI 80 and 5QI 87.

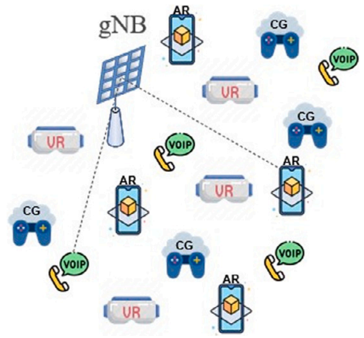


Fig. 1. Single-cell evaluation scenario topology.

Table 1
Single-cell evaluation scenario.

Scenario parameter	Value
Carrier frequency	4 GHz
Bandwidth	150 MHz/75 MHz
Numerology	1
BS transmit power	41 dBm/sector
UE transmit power	23 dBm
BS antenna height	25 m
UE antenna height	1.5 m
Propagation model	3GPP UMa TR 38.901
BS antenna array	1 TXRU: 4×8 (3GPP elements)
BS antenna element gain	8 dBi
BS horizontal/vertical spacing	0.5/0.8 Lambda
UE antenna array	1 TXRU: 1×1 (isotropic element)
UE antenna element gain	0 dBi
BS noise figure	5 dB
UE noise figure	7 dB
Noise power spectral density	−174 dBm/Hz
Min/Max BS-UE distance	2 m/20 m
Shadowing	Disabled
Fading and channel condition updates	No updates
Channel condition	LoS
HARQ Retx	Disabled
DL AR video data rate/FPS	5 Mbps/30 Fps
DL VR video data rate/FPS	30 Mbps/60 Fps
DL CG video data rate/FPS	30 Mbps/60 Fps
Duplexing mode	TDD
TDD pattern	DDDDU
PDCP Discarding Timer	100 ms for 5QI = 1, 10 ms for 5QI = 80, 5 ms for 5QI = 87 [8]

In addition, we have created a simple example in `cttc-nr-simple-qos-sched`, composed of 1 gNB and various UEs (even UEs get voice 5QI = 1 and odd UEs get AR 5QI = 80), to test and validate the correct functionality of the new QoS MAC scheduler. In this example, we can configure the load (full buffer or medium load) of the voice UEs, and see the proper behavior of the scheduler. More precisely, when full buffer of voice UEs is configured, we can see that the resulting throughputs follow the proportion of scheduling priorities ratio.

Finally, for a more extended evaluation of the new QoS MAC scheduler, as presented in this paper, we have created an additional example, known as `cttc-traffic-example-hex.cc` that includes a more complex topology. More details related to the scenario setting and the simulation details are given in the following section.

3.5. Evaluation scenarios

For the evaluation of the new QoS MAC scheduler presented in this paper we study the E2E packet delay and E2E throughput using the example `cttc-traffic-example-hex.cc`, where we consider two different scenarios with mixed traffic (AR/VR/CG/VoIP).

On the one hand, we validate the correct operation of the scheduler under a single-cell scenario composed of 1 gNB and 16 users, where there are 4 AR users, 4 VR users, 4 users play CG, and 4 users are generating VoIP traffic. The single-cell scenario network topology is depicted in Fig. 1. The parameter configuration of the scenario follows the setting of the Dense-eMBB Configuration A, as defined in [23], while in this case we focus on the DL direction assuming good propagation conditions. Table 1 shows the simulation parameters, as well as the configured data rates and frame per seconds (fps) used throughout the simulations. Notice

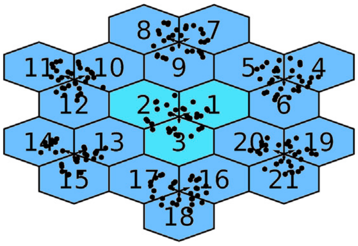
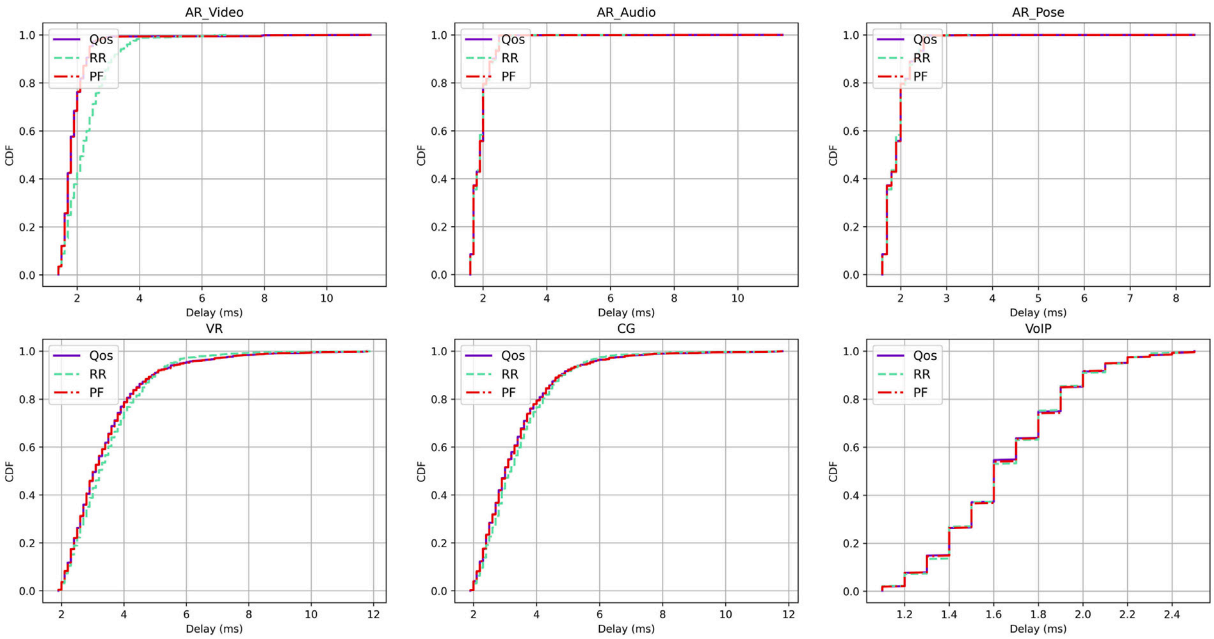


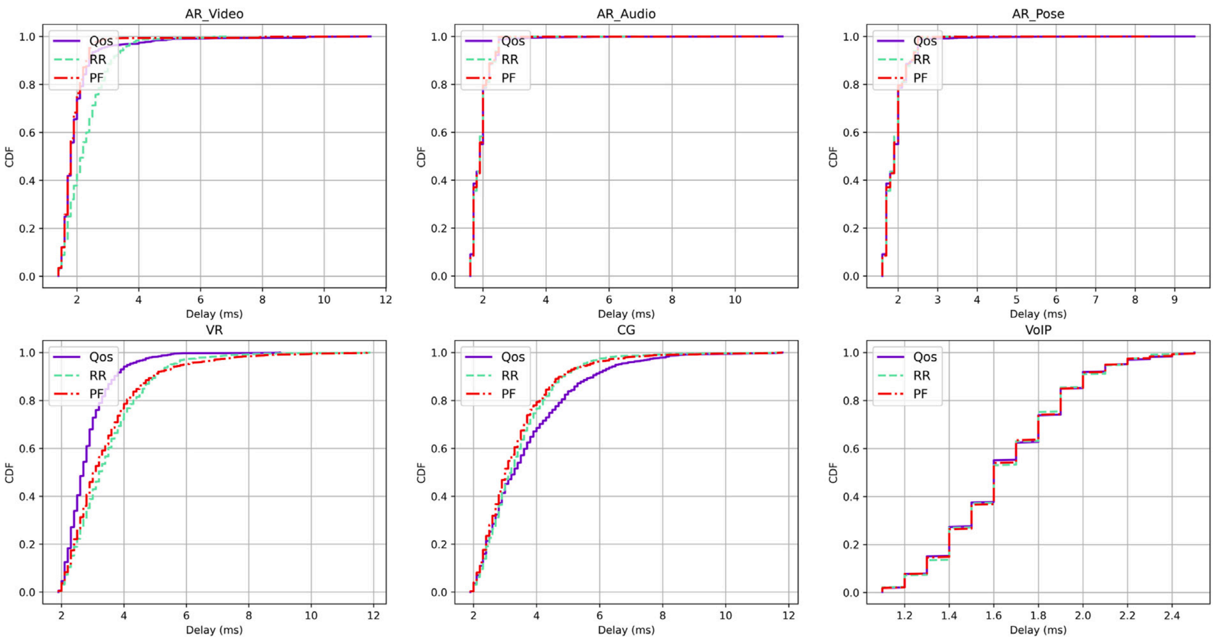
Fig. 2. Rural-eMBB evaluation scenario topology.

Table 2
Rural-eMBB configuration A evaluation scenario.

Scenario parameter	Value
Carrier frequency	700 MHz
Bandwidth	10 MHz
Inter-site distance	1732 m
Sectors	30/150/270 degrees
Numerology	1
BS transmit power	46 dBm/sector
UE transmit power	23 dBm
BS antenna height	35 m
UE antenna height	1.5 m
Propagation model	3GPP RMa TR 38.901
BS antenna array	1 TXRU: 8 × 1 (3GPP elements)
BS antenna element gain	8 dBi
BS horizontal/vertical antenna spacing	0.5/0.8 Lambda
UE antenna array	1 TXRU: 1 × 1 (isotropic element)
UE antenna element gain	0 dBi
BS noise figure	5 dB
UE noise figure	7 dB
Noise power spectral density	−174 dBm/Hz
Min/Max BS-UE distance	10 m/500 m
Shadowing	Enabled
Fading and channel condition updates	20 ms /100 ms
Channel condition	LoS/NLoS (stochastic model)
HARQ Retx	Enabled
DL AR video data rate/FPS	1 Mbps/30 Fps
DL VR video data rate/FPS	5 Mbps/60 Fps
DL CG video data rate/FPS	5 Mbps/60 Fps
UL AR video data rate/FPS	1 Mbps/30 Fps
Duplexing mode	TDD
TDD pattern	DDDDU
PDCP Discarding Timer	100 ms for 5QI = 1, 10 ms for 5QI = 80, 5 ms for 5QI = 87 [8]



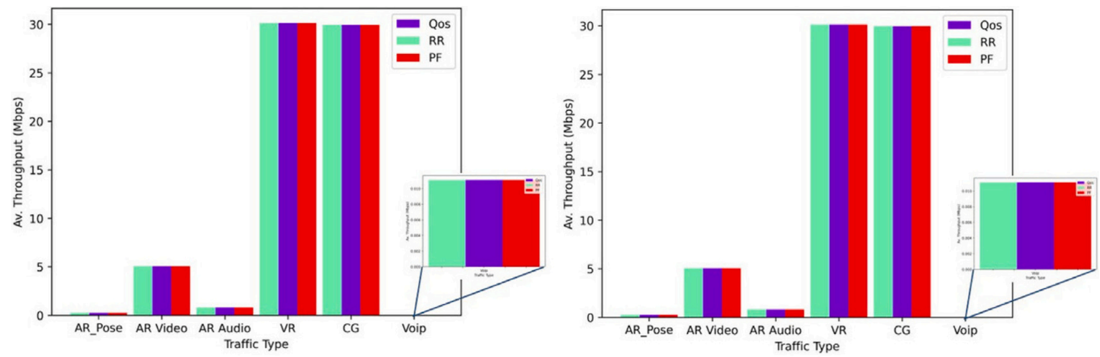
(a) *5QI80 for all XR*



(b) *5QI80 for AR/CG and 5QI87 for VR*

Fig. 3. Packet delay CDFs for non-saturated study case.

that for the validation process we study the single-cell scenario under two different setups and for this reason we vary the defined bandwidth. In the first case we set a high bandwidth and we study a *non-saturated* situation, meaning that the available bandwidth, and consequently the available resources, are enough to serve all the traffic generated by all users of the network. This use case is the most common approach for the validation of a scheduler. However, since in such case all the traffic is served, the throughput cannot give us information related to the user prioritization, user fairness, and user starvation. Moreover, the closer the system is to



(a) *5QI80 for all XR* (b) *5QI80 for AR/CG and 5QI87 for VR*

Fig. 4. E2E throughput for non-saturated study case.

a saturation mode, the more critical becomes the role of the QoS scheduler. For this reason, we have also opted to study a *saturated* case, where the bandwidth is decreased and therefore, the available resources are not enough to serve all the traffic.

On the other hand, we study a multi-cell scenario with hexagonal topology consisting of 21 gNBs (7 sites with 3 sectors per site). In each sector, 10 UEs are randomly deployed following a uniform distribution. The network topology of the multi-cell scenario is depicted in Fig. 2. Notice that the light blue color corresponds to the inner ring, composed by 1 site with 3 gNBs, while the darker blue depicts the outer ring that has 6 more sites with 3 gNBs each, resulting in a total of 21 gNBs in the whole scenario. The black dots depict the UEs deployed in the network. Out of 10 UEs for each gNB, 2 UEs carry AR traffic, 2 UEs carry VR traffic, 2 UEs play CG and 4 UEs generate VoIP traffic. The scenario parameters are as defined for the Rural-eMBB Configuration A scenario in [23]. In Table 2 we present a summary of them. Notice that in this case we study both DL and UL directions.

4. Simulation results

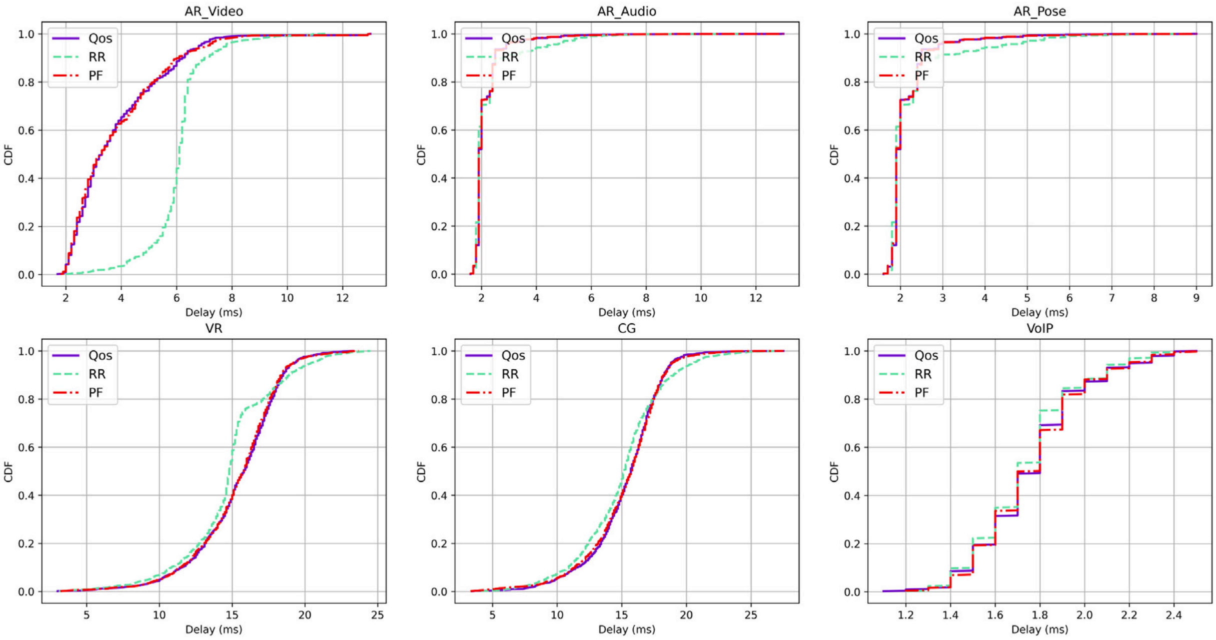
For the validation and evaluation of the proposed QoS MAC scheduler for mixed traffic we consider different 5QIs for the different types of traffic, with target to study the behavior of the scheduler with respect to the E2E packet delay and E2E throughput per traffic flow type. We compare the proposed scheduler against traditional schedulers, such as the RR and PF. The results are presented in the next two subsections for the single-cell and multi-cell scenarios.

4.1. Single-cell validation scenario

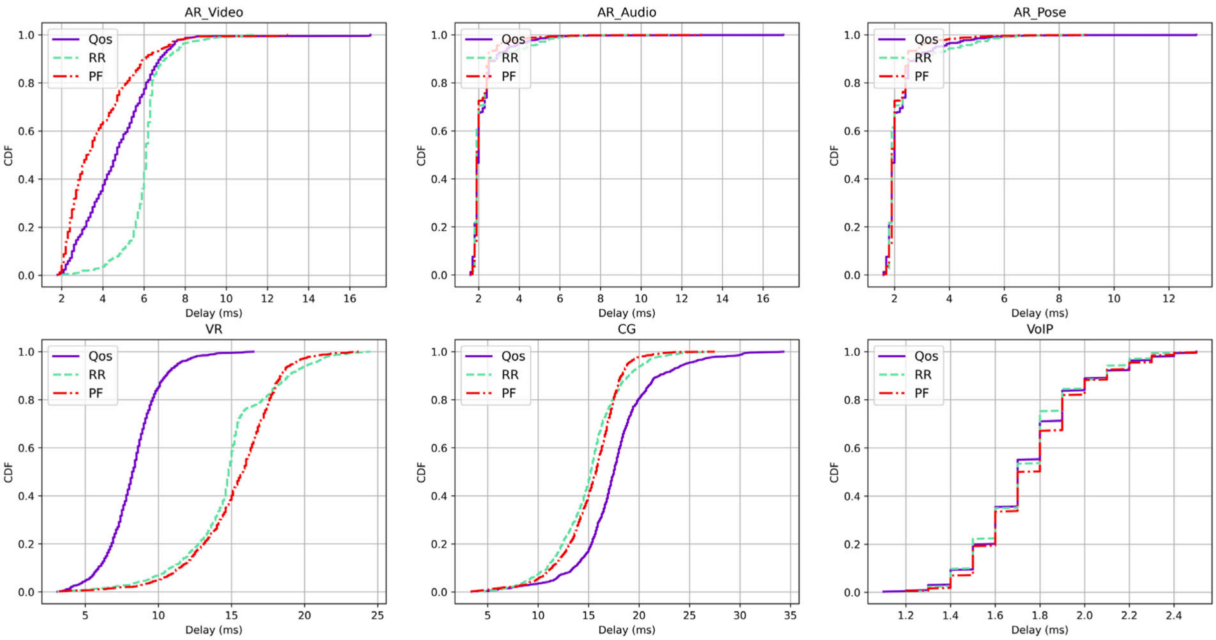
As a first step, we target the validation of the proper functionality of the presented QoS MAC scheduler though a single-cell scenario with good propagation conditions as described in Section 3.5. Out of the 16 UEs deployed in the network, 4 carry AR traffic composed of 3 flows (video, audio-data and pose-control), 4 UEs generate VR traffic of 1 flow, 4 UES have CG traffic of 1 flow and 4 UEs generate VoIP traffic (see Fig. 1). For the validation we study two cases with respect to the assigned 5QIs and two cases for the traffic load. To do so, we consider one case where all XR UEs have the same 5QI = 80 (AR, VR and CG UEs) and the VoIP UEs have 5QI = 1 (let us refer to it as *5QI80 for all XR* case), and another case where the VR UEs have 5QI equal to 87, while the rest of XR (i.e., AR and CG) have 5QI = 80 and the VoIP 5QI = 1 (let us refer to it as *5QI80 for AR/CG and 5QI87 for VR* case). With this differentiation we aim to show the behavior of the benchmark schedulers and the new QoS MAC scheduler with respect to the treatment of the traffics based on the resulting E2E packet delay and throughput. Moreover, we study both non-saturated and saturated situations, as presented in Section 3.5, therefore we evaluate for two different bandwidths. For the non-saturated case we set the bandwidth to 150 MHz and for the saturated case we decrease the bandwidth to 75 MHz with target to study the performance of the schedulers in the case that there are not enough resources to serve all the generated traffic.

Fig. 3 presents the Cumulative Distribution Functions (CDFs) of the packet delay for each traffic flow for the case of bandwidth of 150 MHz (non-saturation), comparing the proposed QoS MAC scheduler performance against the transitional RR and PF schedulers and studying the *5QI80 for all XR* (top figure) versus *5QI80 for AR/CG and 5QI87 for VR* (bottom figure) cases.

As it can be seen from the figures, for both cases, since there is enough bandwidth to serve all the generated traffic from all applications, all schedulers perform similarly, however for the case that VR has been assigned a 5QI with higher priority (5QI = 87), we can clearly observe that the new QoS MAC scheduler proposed in this paper prioritizes this type of traffic, at the cost of course of a small increment in the delay of lower priority traffic (i.e. CG traffic). In particular, the proposed scheduler presents an average reduction of 17% in the delay of VR traffic with respect to PF, at the cost of an average increment of 10% in the delay of CG. This confirms the correct functionality of the proposed QoS MAC scheduler, and in addition demonstrates that it can manage correctly the traffic flows based on the desired priority.



(a) *5QI80 for all XR*

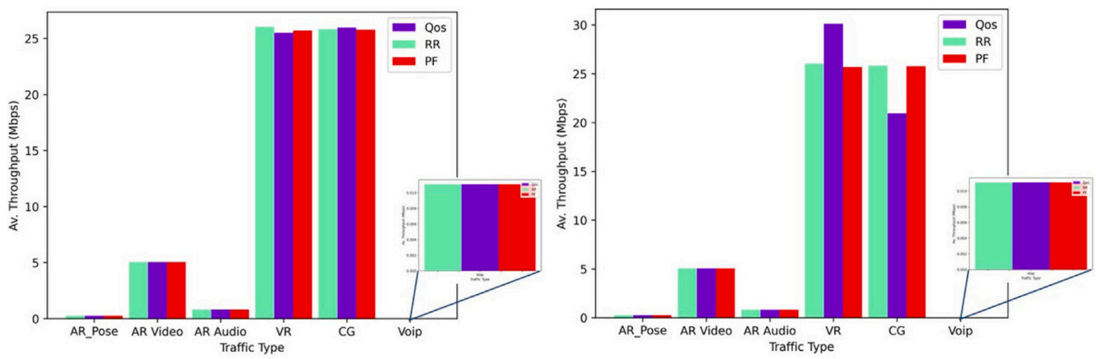


(b) *5QI80 for AR/CG and 5QI87 for VR*

Fig. 5. Packet delay CDFs for saturated study case.

Moreover, Fig. 4 presents the resulted E2E throughput for the *5QI80 for all XR* (left figure) versus *5QI80 for AR/CG and 5QI87 for VR* (right figure) cases. As it can be seen, all traffics are served and the targeted throughput is reached (see Table 1 data rates) for both cases and all schedulers, considering that the network is not saturated.

In addition, we present the study of the scheduler's behavior in a saturated scenario, where the available bandwidth has been reduced to 75 MHz. Fig. 5 presents the packet delay CDFs for each traffic flow, comparing the QoS MAC scheduler against the



(a) *5QI80 for all XR* (b) *5QI80 for AR/CG and 5QI87 for VR*

Fig. 6. E2E Throughput for saturated study case.

benchmark RR and PF, while Fig. 6 depicts the achieved E2E throughput. Again we evaluate the performance of the schedulers under the *5QI80 for all XR* and *5QI80 for AR/CG and 5QI87 for VR* cases.

Focusing on the case of *5QI80 for all XR*, one can see that for both packet delay and throughput, all schedulers present similar performance, while from the throughput results, it is obvious that the network is in saturation and not all the generated traffic can be served (e.g., see VR and CG traffic). On the other hand, when the 5QI of the VR traffic is set to 87 (i.e., higher priority), it is clearly seen that the proposed QoS MAC scheduler presents a significant packet delay reduction and throughput increment for this type of traffic, at a small cost on the delay of AR Video and CG traffic types, and on the throughput of CG. In particular, compared to PF, the QoS scheduler presents a reduction of 46% for the average VR traffic delay, while the average AR Video and CG delay increments are of the order of 24% and 16%, respectively. As such, it is demonstrated that the proposed QoS MAC scheduler is able to treat the various traffic types as desired, based on the traffic QoS Flow Profile.

4.2. Multi-cell calibrated scenario

In order to have a complete evaluation of the presented QoS MAC scheduler, we consider a realistic scenario as described in Section 3.5, to study the performance under real-network conditions. In this scenario we study both DL and UL directions, where for each gNB out of 10 UEs, 2 UEs carry DL AR traffic of 3 flows and UL AR of 1 flow, 2 UEs generate DL VR traffic of 1 flow, 2 UEs have DL CG traffic of 1 flow and 4 UEs carry DL and UL VoIP traffic of 1 flow each. For this scenario, we consider only the *5QI80 for AR/CG and 5QI87 for VR* case, since it is the one of higher interest. The parameter setting is as presented in Table 2. It is worth noting that we study a saturation case, since the bandwidth defined in this scenario, along with the network interference and the XR applications data rates, result in a saturation state for the downlink.

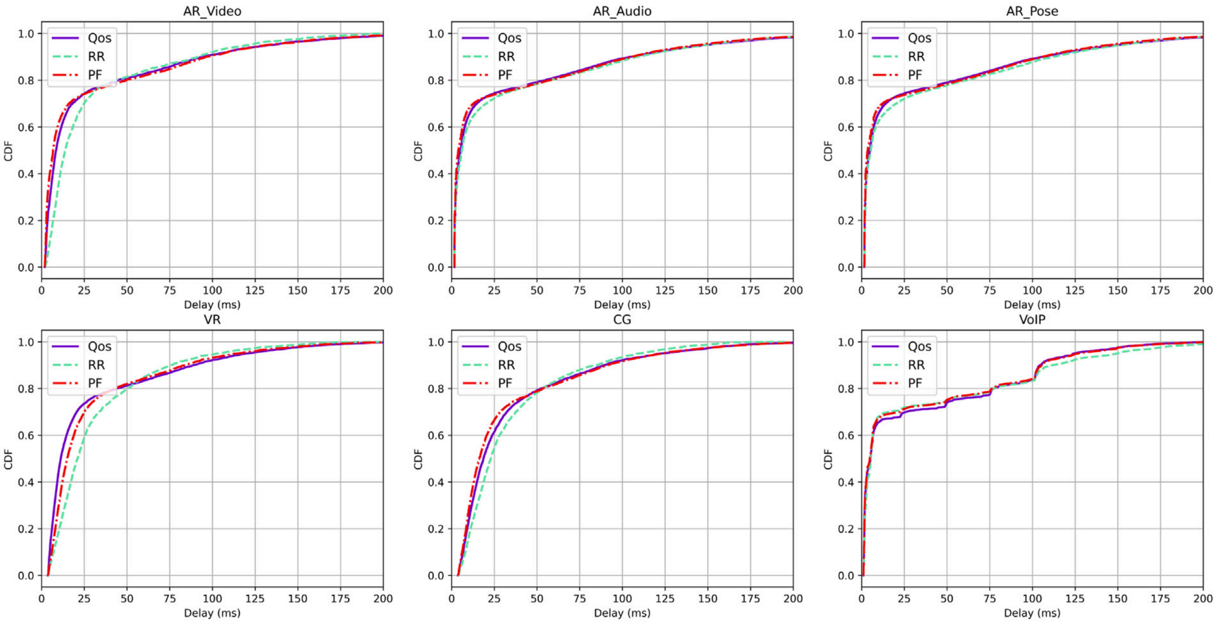
Fig. 7 presents the packet delay CDFs for the Rural-eMBB scenario. Figure (a) shows the results of the packet delay for the *5QI80 for AR/CG and 5QI87 for VR* for saturation case in the DL direction, while Figure (b) depicts the results for the UL case. In addition, Fig. 8 (a) presents the E2E throughput for the DL direction and Fig. 8 (b) for the UL direction.

As it can be observed, in the DL direction the proposed QoS MAC scheduler clearly prioritizes the VR traffic presenting an average delay reduction of 4.2%, demonstrating that it is capable to classify the traffic flows based on the QoS indicators even in scenarios with very high interference and in saturated network condition. With respect to the UL direction, it can be seen that the proposed solution performs similarly to the traditional schedulers demonstrating the correct operation also in the case of UL traffics (notice that the UL is not saturated).

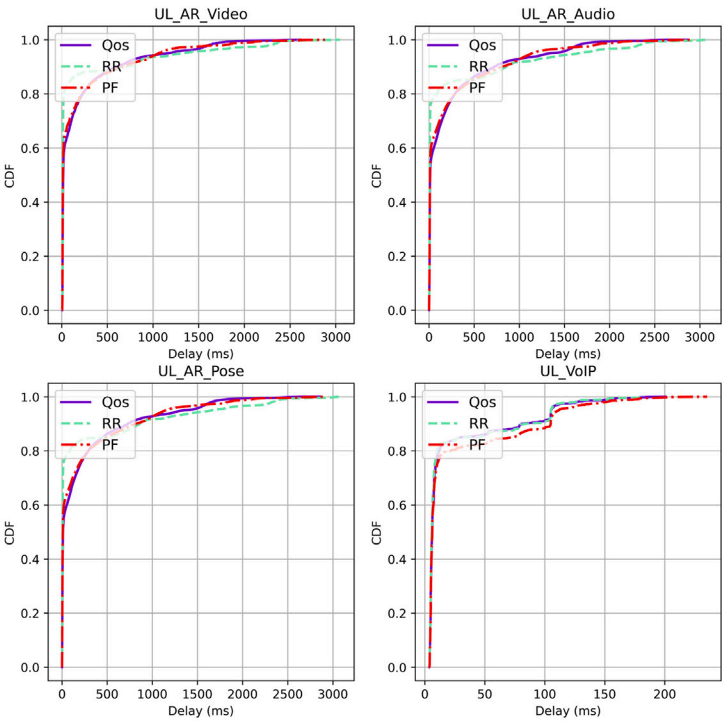
5. Conclusions

In this paper, we have presented an E2E system-level evaluation of QoS schedulers for XR traffic in 5G NR mixed traffic scenarios. In particular, we have proposed, validated and evaluated a new QoS MAC scheduler that exploits the QoS model introduced in 5G NR that is especially beneficial for XR traffic applications. We have used the 5G-LENA simulator, recently calibrated under 3GPP reference scenarios, in order to validate the correct operation of the scheduler in a single-cell scenario of mixed traffic, including AR/VR/CG and VoIP. Moreover, we have carried out the performance evaluation of the proposed solution in a 3GPP multi-cell reference scenario used for the calibration of the 5G-LENA simulator. Results have been compared against traditional schedulers and have demonstrated that the proposed QoS MAC scheduler is able to treat the various traffic flows in accordance to the QoS indicators, such as the traffic priority level.

This work is a preliminary study for more complex traffic scenarios and paves the way for various scheduling approaches to be evaluated. In this direction, future work is envisioned to perform E2E system-level evaluations of the proposed QoS MAC scheduler for XR traffic that employs other QoS indicators, such as the resource type (GBR/delay-critical-GBR/non-GBR) and the packet delay



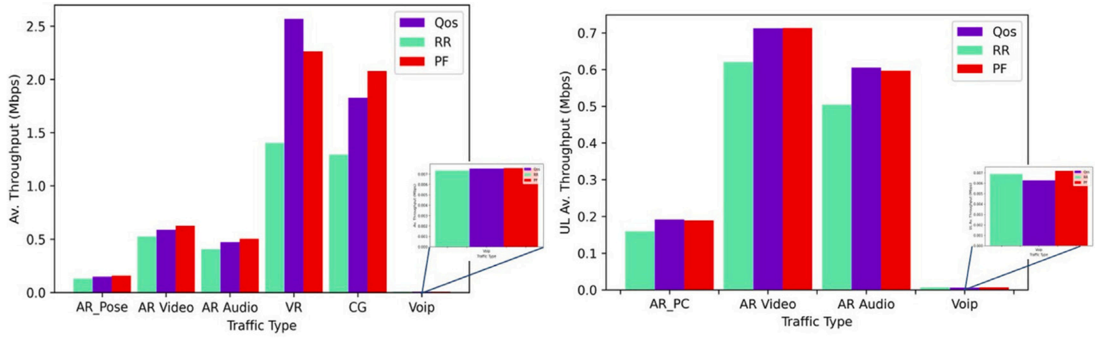
(a) *DL Direction*



(b) *UL Direction*

Fig. 7. Packet delay CDFs for Rural-eMBB study case.

budget, as well as to consider additional design options for the calculation of the scheduling weight in case of multi-QoS flows per user with flow differentiation. In this last case it would be really interesting to include also a comparison of the 1x1 and Mx1 architectures.



(a) *DL Direction* (b) *UL Direction*

Fig. 8. E2E Throughput for Rural-eMBB study case.

Data availability

The code is open-source.

Acknowledgments

CTTC authors have received funding by TSI-063000-2021-56/57 (6G-BLUR project) from the Spanish Government, Grant PID2021-126431OB-I00 funded by MCIN/AEI/10.13039/501100011033 and by "ERDF A way of making Europe", and Generalitat de Catalunya, Spain grant 2021 SGR 00770.

References

[1] Huawei, Empowering consumer-focused immersive VR and AR experiences with mobile broadband, 2016, Available at https://www.huawei.com/en/technology-insights/industry-insights/outlook/mobile-broadband/insights-reports/vr-and-ar.
[2] M. Lecci, M. Drago, A. Zanella, M. Zorzi, An open framework for analyzing and modeling XR network traffic, IEEE Access 9 (2021) 129782–129795, http://dx.doi.org/10.1109/ACCESS.2021.3113162.
[3] The Goldman Sachs Group, Inc., Virtual & augmented reality: Understanding the race for the next computing platform, 2016, Available at https://www.goldmansachs.com/insights/pages/virtual-and-augmented-reality-report.html.
[4] Qualcomm, The mobile future of extended reality (XR), 2020, Available at https://www.qualcomm.com/research/extended-reality.
[5] Ericsson, XR and 5G: Extended reality at scale with time-critical communication, 2021, Available at https://www.ericsson.com/en/reports-and-papers/ericsson-technology-review/articles/xr-and-5g-extended-reality-at-scale-with-time-critical-communication.
[6] Revised SI on XR evaluations for XR, in: Qualcomm, RP-201145, 3GPP TSG-RAN 88e Meeting, 2020.
[7] 3GPP, TR 38.838, study on XR (extended reality) evaluations for NR, V17.0.0, 2022.
[8] TSG ssa; system architecture for the 5G system; stage 2, 3GPP TS 23.501, 2017, Release 15, v15.0.0.
[9] L. Miuccio, D. Panno, P. Pisacane, S. Riolo, A QoS-aware and channel-aware radio resource management framework for multi-numerology systems, Comput. Commun. 191 (2022) 299–314.
[10] Z. Gu, C. She, W. Hardjawana, S. Lumb, D. McKechnie, T. Essery, B. Vucetic, Knowledge-assisted deep reinforcement learning in 5G scheduler design: From theoretical framework to implementation, IEEE J. Sel. Areas Commun. 39 (7) (2021) 2014–2028.
[11] S. Barmpounakis, N. Maroulis, N. Koursioumpas, A. Kousaridas, A. Kalamari, P. Kontopoulos, N. Alonistioti, AI-driven, QoS prediction for V2X communications in beyond 5G systems, Comput. Netw. 217 (2022) 109341.
[12] A. Mamane, M. Fattah, M.E. Ghazi, M.E. Bekkali, Y. Balboul, S. Mazer, Scheduling algorithms for 5G networks and beyond: Classification and survey, IEEE Access 10 (2022) 51643–51661.
[13] M. Mahfoudi, M.E. Bekkali, A. Najd, M. El Ghazi, S. Mazer, A new downlink scheduling algorithm proposed for real time traffic in LTE system, Int. J. Electron. Telecommun. vol. 61 (4) (2015).
[14] G. Piro, L.A. Grieco, G. Boggia, F. Capozzi, P. Camarda, Simulating LTE cellular systems: An open-source framework, IEEE Trans. Veh. Technol. 60 (2) (2011) 498–513.
[15] A.M. Nor, O. Fratu, S. Halunga, A. Alyosef, Z.D. Zaharis, S. Rizou, P.I. Lazaridis, Demand based proportional fairness scheduling for 5G eMBB services, in: 2022 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom), 2022, pp. 263–268.
[16] I.-S. Comsa, R. Trestian, G.-M. Muntean, G. Ghinea, 5MART: A 5G SMART scheduling framework for optimizing QoS through reinforcement learning, IEEE Trans. Netw. Serv. Manag. 17 (2) (2020) 1110–1124.
[17] A. Kaur, K. Kumar, A reinforcement learning-based green resource allocation for heterogeneous services in cooperative cognitive radio networks, IEEE Trans. Netw. Serv. Manag. 19 (2) (2022) 1554–1566.
[18] 5G-LENA, 2022, Available at https://5g-lena.cttc.es/.
[19] K. Koutlia, B. Bojovic, Z. Ali, S. Lagén, Calibration of the 5G-LENA system level simulator in 3GPP reference scenarios, Simul. Model. Pract. Theory 119 (2022) 102580.
[20] T.R. Henderson, M. Lacage, G.F. Riley, C. Dowell, J. Kopena, Network simulations with the ns-3 simulator, SIGCOMM Demonstr. 14 (2008).
[21] N. Patriciello, S. Lagen, B. Bojovic, L. Giupponi, An E2E simulator for 5G NR networks, Simul. Model. Pract. Theory 96 (2019) 101933, http://dx.doi.org/10.1016/j.simpat.2019.101933.
[22] 3GPP, Study on Channel Model for Frequencies from 0.5 to 100 GHz, TR 38.901 (Rel. 15), v16.1.0, 2020.
[23] ITU-R, Submission, Evaluation Process and Consensus Building for IMT-2020, ITU-R IMT-2020/2-E, 2019.

[24] N. Patriciello, S. Lagen, L. Giupponi, B. Bojovic, An improved MAC layer for the 5G NR Ns-3 module, in: Proceedings of the 2019 Workshop on Ns-3, in: WNS3 '19, Association for Computing Machinery, New York, NY, USA, 2019, pp. 41–48.
[25] Nr module manual, 2022, Available at https://cttc-lena.gitlab.io/nr/nrmodule.pdf.
[26] TSG ran; nr; NR and NG-RAN overall description; stage 2, 3gpp ts 38.300, 2022, Release 17, v17.0.0.
[27] 5G Americas, 5G services innovation, 2019, Available at https://www.5gamericas.org/wp-content/uploads/2019/11/5G-Services-Innovation-FINAL-1.pdf.