

RESEARCH

Open Access



# 3GPP QoS-based scheduling framework for LTE

Pablo Ameigeiras<sup>1\*</sup>, Jorge Navarro-Ortiz<sup>1</sup>, Pilar Andres-Maldonado<sup>1</sup>, Juan M. Lopez-Soler<sup>1</sup>, Javier Lorca<sup>2</sup>, Quiliano Perez-Tarrero<sup>3</sup> and Raquel Garcia-Perez<sup>3</sup>

## Abstract

This paper proposes the design of a scheduling framework for the downlink of the Long Term Evolution (LTE) system with the objective of meeting the Quality of Service (QoS) requirements as defined by the QoS architecture of the 3G Partnership Project (3GPP) specifications. We carry out a thorough review of 3GPP specifications analyzing the requirements of the 3GPP QoS architecture. LTE bearers may be associated with a Guaranteed Bit Rate (i.e., GBR bearers) or not (i.e., non-GBR bearers). Additionally, the specifications establish a Packet Delay Budget (PDB) to limit the maximum packet transfer delay. To achieve our goal, we design a channel-aware service discipline for GBR bearers which is able to fulfill not only the GBR but also the PDB. Additionally, we also design an algorithm for prioritizing GBR and non-GBR bearers from different QoS Class Identifiers (QCIs) following 3GPP QoS rules. We compare the proposed framework with two reference schedulers by means of network-level simulations. The results will show the ability of the proposed framework to address the QoS requirements from 3GPP specifications while providing an interesting performance from a spectral efficiency viewpoint.

**Keywords:** LTE, Scheduling, GBR, Non-GBR, QoS Class Identifier, PDB

## 1 Introduction

Numerous operators worldwide have commercially launched Long Term Evolution (LTE) and LTE-Advanced (LTE-A) networks in recent years. The 3G Partnership Project (3GPP) has defined these radio access technologies with the objective of providing improved network capacity, coverage, and latency. LTE is based on Orthogonal Frequency Division Multiple Access (OFDMA) where the resource blocks are assigned by an eNodeB scheduler, which therefore plays a key role in the system performance. LTE-A also includes various features that enhance LTE performance.

3GPP specifications have also standardized a Quality of Service (QoS) architecture [1] different from the one defined for previous 3G radio access networks. This architecture is based on the fundamental concept of the bearer, which is assigned to one predefined QoS class. This class-based association determines the final QoS attributes of

the provided services to the subscriber groups. The architecture defines relative priorities for the QoS classes, and depending on its class, the bearer is associated with a Guaranteed Bit Rate (i.e., GBR bearer) or not (i.e., non-GBR bearer).

The objective of our investigation is the design of a scheduling framework for the downlink of the LTE system which satisfies the QoS requirements defined by the QoS architecture of 3GPP specifications [1, 2]. To the best of the authors' knowledge, no previous work has achieved that objective. For that purpose, we will review the QoS concept in 3GPP specifications. We will show that 3GPP specifications establish a per-QoS-class upper bound for the delay of the data packets transferred by a bearer. This bound is named Packet Delay Budget (PDB) by the 3GPP specifications, and they establish it as the primary goal of the scheduling framework. We will review the terms indicated in 3GPP specifications regarding the satisfaction of the PDB for GBR and non-GBR bearers.

For these reasons, we propose in this paper an innovative scheduling framework for the downlink of LTE. The novelty of our design is that it aims at globally addressing the QoS requirements as defined by 3GPP specifications.

\*Correspondence: pameigeiras@ugr.es

<sup>1</sup>Department of Signal Theory, Telematics and Communications, University of Granada, Periodista Daniel Saucedo Aranda s/n, 18071 Granada, Spain  
Full list of author information is available at the end of the article

For this, we carry out the following contributions: (i) a thorough review of 3GPP specifications analyzing the QoS requirements imposed by the 3GPP QoS architecture and their implications on the scheduling design. From this review, we identify the Packet Delay Budget as a key requirement to be fulfilled. Additionally, we identify that if the PDB cannot be fulfilled for all bearers then prioritization between bearers of different QoS classes should be triggered. (ii) The design of a channel-aware service discipline for GBR bearers that is able to fulfill not only the Guaranteed Bit Rate but also the Packet Delay Budget. The discipline incorporates a delay-dependent factor based on a sigmoid function. The benefit of this delay-dependent factor is that a parameter controls its upper bound. This facilitates the prioritization of bearers of different QoS classes compared to other service disciplines. (iii) The design of an algorithm for prioritizing GBR and non-GBR bearers from different QoS Class Identifiers (QCIs) following 3GPP QoS rules when the PDB cannot be met for all bearers.

We have evaluated the proposed framework by means of network-level simulations and compared it with two reference schedulers. For the evaluation, we have considered scenarios with different load levels and traffic mixes of real-time, progressive video, and elastic traffic. The results will show the ability of the proposed framework to satisfy the QoS requirements from 3GPP specifications while providing an interesting performance from a spectral efficiency viewpoint.

The rest of the paper is organized as follows: Section 2 provides an overview of related works. Section 3 presents the system model. Section 4 describes the QoS concept in 3GPP and discusses its implications on the scheduling design. Section 5 presents the proposed QoS scheduling design and Section 6 its performance results by means of simulations. Section 7 finally draws the main conclusions.

## 2 Related works

The literature on downlink scheduling for LTE and OFDMA systems is extensive. Sadiq et al. [3], Capozzi et al. [4], and Dardouri and Bouallegue [5] provide interesting overviews of related prior work.

### 2.1 QoS-aware strategies for real-time traffic

Concentrating on QoS-aware scheduling algorithms, the Modified-Largest Weighted Delay First (M-LWDF) [6], the Exponential/Proportional Fair (EXP/PF), the Exp rule, and the Log rule are relevant proposals described in those overviews. Despite these scheduling algorithms providing very interesting performance benefits [3–5], they are not well suited to fulfilling the QoS requirements of 3GPP specifications. They were originally designed for a scenario with real-time traffic only. Therefore, they require to be extended to support non-real-time traffic and a

strategy to provide relative prioritization between traffic classes. Moreover, although these strategies increase the user's priority when the Head of Line (HOL) packet delay increases, they may be enhanced by emphatically increasing the user's priority when the HOL packet delay approaches its upper bound. The urgency of the HOL packet delay is addressed in the Delay-Prioritized Scheduling (DPS) algorithm [7] by prioritizing a user according to  $\delta = D - w$ , where  $w$  denotes the user's HOL packet delay and  $D$  the packet delay upper bound. However, DPS has also been designed for a real-time traffic scenario only.

### 2.2 QoS-aware strategies for heterogeneous traffic

Some other algorithms have been developed for a scenario with a mix of real-time and non-real-time traffic. The Rate-Level-Based Scheduling (RLBS) [8] algorithm prioritizes a user according to their  $\delta$ , their spectral efficiency, and a GBR-related factor if the user supports a GBR bearer. However, it is not appropriate to apply a packet delay upper bound to non-real-time traffic that is delay tolerant (see Section 4.3). In [9], Ai et al. propose to compute the service order of real-time users based on the factor  $D - \beta \cdot w$ , where  $\beta$  is an adaptive delay adjustment. In the last step, the algorithm assigns the remaining resource blocks to the non-real-time users. In [10], Iturralde et al. propose to implement a virtual token mechanism along with either the M-LWDF or the EXP/PF scheduling algorithms for real-time traffic. Their strategy provides very interesting performance for real-time flows, but it penalizes non-real-time flows. In [11], Nasralla and Martini suggest a modification to Iturralde's proposal by incorporating the queue size and the HOL packet delay in the priority computation for both real-time and non-real-time traffic. Their results show a balanced performance between the flows of the different traffic classes. In [12], Capozzi et al. propose the Frame Label Scheduler (FLS), which is composed of two levels. At the highest level, FLS calculates every frame (i.e., 10 ms) the total amount of data that real-time flows should transmit in the following frame in order to satisfy their delay constraints. At the lowest level, every Transmission Time Interval (TTI) (i.e., 1 ms) FLS assigns resource blocks to each real-time flow following a Maximum Throughput policy. Then, Proportional Fair is used to share the spare spectrum among best effort users.

### 2.3 QoS-aware strategies for bit rate guarantees

Other algorithms have focused on providing a guaranteed bit rate. In [13], Mongha et al. present a decoupled time/frequency domain scheduler intended for providing a target bit rate to GBR bearers and fairness control to non-GBR bearers. The proposal of Zaki et al. in [14] is based on the previous one, but it includes a relative

prioritization between the bearers of the different QoS classes. Mongha's and Zaki's proposals do not increase the user's priority when the HOL packet delay approaches its upper bound, and therefore, they are not well suited to satisfy the PDB. In [15], Gora proposes a QoS-aware resource management for LTE-A-based relay networks. Although his proposal concentrates on multi-hop relays, it presents valuable utility functions to satisfy the GBR and/or the PDB for real-time traffic. However, his proposal has been neither designed nor evaluated in a traffic mix scenario with several QCIs.

All these proposals presented in Sections 2.2 and 2.3 are not well suited to fulfilling the QoS requirements of 3GPP specifications. They either do not provide relative prioritization between QoS classes or they provide relative prioritization without following the rules imposed by 3GPP specifications. Furthermore, the large majority of these related works do not include users' birth and death processes in their evaluation models, which may strongly impact their obtained results (see [16]).

### 3 System model

Let us consider the OFDMA downlink transmission of an LTE cell where a base station transmits data to a set of users. Let  $K = \{1, \dots, k, \dots, K\}$  represent both the set of users and its cardinality at an arbitrary epoch in time. The notation used in the paper is given in Table 1. The base station carries out the transmission towards the  $K$  users in TTIs of fixed duration  $T = 1$  ms. The total available bandwidth for transmission is divided in  $S$  resource blocks of fixed size. Each resource block is composed of 12 consecutive subcarriers with a subcarrier spacing of 15 kHz.

The base station transmission power is assumed to be constant across all subcarriers. The channel model assumes small-scale fading, shadow fading, and path loss. The signal received by each user is corrupted with AWGN noise and intercell interference. The resulting signal-to-interference ratio  $\text{SINR}_k[n, s]$  of user  $k$  on resource block  $s$  and TTI  $n$  is used to compute its achievable transmission rate  $R_k[n, s]$  based on a finite set of Modulations and Coding Schemes (MCSs). The base station is assumed to have instantaneous and perfect knowledge of the achievable transmission rate  $R_k[n, s]$  of all the users based on the Channel Quality Indicator (CQI) reports. The scheduling algorithm uses this information to assign bandwidth resources to the users. The granularity of the bandwidth assignment equals one resource block:

$$\Psi_k[n, s] = \begin{cases} 1 & \text{resource block } s \text{ is assigned to user } k \text{ in TTI } n \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

The base station applies link adaptation, and it employs a single MCS for all resource blocks assigned to user  $k$  in TTI  $n$ . If user  $k$  successfully decodes the radio block

**Table 1** Notation used in this paper

$k, K$	Bearer index and set of all bearers in cell
$s, S$	Resource block index and number of available resource blocks
$m, M$	QCI index and set of all QCIs
$T$	Transmission Time Interval
$P_k[n, s]$	Priority of bearer $k$ on resource block $s$ and TTI $n$
$R_k[n, s]$	Achievable transmission rate of bearer $k$ on resource block $s$ and TTI $n$
$R_{\text{MCS}}$	Rate of selected MCS
$\overline{r}_k[n]$	Low pass filtered data rate that bearer $k$ has received until TTI $n$
$\alpha$	Degree of fairness factor
$w_k[n]$	Head of Line delay of bearer $k$ measured in TTI $n$
$q_k[n]$	Number of bits in queue of bearer $k$ measured in TTI $n$
$f(w_k)$	Delay-dependent factor in priority computation
$D$	Target packet delay of delay-dependent factor
$a_k$	Slope adjustment factor for sigmoid function
$r_{lb}$	Lower bound for $\overline{r}_k[n]$
$F_k^{\text{QCI}_m}$	QCI relative prioritization factor for bearer $k$ mapped onto $\text{QCI}_m$
$Q_k$	Quality performance indicator for bearer $k$ , (see (9))
$q_k[n]$	Queue length of bearer $k$ in TTI $n$
$\lambda_k$	Estimator of the average arrival bit rate on bearer $k$
$r_k[n]$	Transmitted data rate in TTI $n$ by bearer $k$
$Q^{\text{QCI}_m}$	Quality performance indicator of the bearer with worst quality among all bearers mapped onto $\text{QCI}_m$ that have sufficient radio channel quality
$TQ^{\text{QCI}_m}$	Minimum quality level for bearer $k$ mapped on to $\text{QCI}_m$
$F$	Priority enhancing factor

in TTI  $n$ , he is assumed to correctly receive  $R_{\text{MCS}} \cdot T$  bits of data for each resource block  $s$  of the radio block.  $R_{\text{MCS}}$  denotes the rate of the selected MCS. The base station has one queue buffer for each user. Let  $q_k[n]$  denote the number of bits of all packets stored in the queue of user  $k$ . Whenever the base station receives a packet destined for user  $k$ , it stores the packet in the user's queue buffer and it updates  $q_k[n]$ . Additionally, it starts a timer for the received packet. This timer is used to obtain the HOL packet delay  $w_k[n]$  when the packet reaches the front of the queue. If user  $k$  successfully decodes the radio block in TTI  $n$ , the queue  $q_k[n]$  corresponding to the flow of user  $k$  is decremented according to the transmitted bits:

$$q_k[n] = q_k[n-1] - \sum_{s=1}^S R_{\text{MCS}} \cdot T \cdot \Psi_k[n, s] \quad (2)$$

Every user in set  $K$  is assigned one bearer to transport its data packets, and let  $k$  denote not only the user index

but also the index of its assigned bearer. Let us also denote  $M = \{1, \dots, m, \dots, M\}$  to represent both the set of QCIs and its cardinality. The assignment of bearer  $k$  to  $QCI_m$  will be denoted as  $k \in QCI_m$ .

## 4 QoS concept in 3GPP

### 4.1 The EPS bearer concept

The Evolved Packet System (EPS) bearer is the basis of the QoS in LTE [1]. It provides a logical channel between the user entity (UE) and a Packet Data Network (PDN) for transporting IP traffic [17]. The EPS bearer is the degree of granularity for bearer-level QoS control. All packets transferred by an EPS bearer are equally treated by forwarding functionalities (e.g., scheduling policy and queue management policy) [1]. EPS bearers can be classified as GBR or non-GBR bearers. An EPS bearer is named a GBR bearer if the EPS permanently allocates dedicated network resources (more specifically, a GBR value) to it. Otherwise, it is named a non-GBR bearer.

The 3GPP QoS concept includes four per-EPS-bearer parameters, three of which are relevant for scheduling purposes [1]:

- *QoS Class Identifier (QCI)*: is a reference scalar used to specify node-specific parameters that control bearer-level packet forwarding treatment.
- *Guaranteed Bit Rate (GBR)*: is the bit rate the GBR bearer is expected to provide.
- *Maximum Bit Rate (MBR)*: is the maximum bit rate a GBR bearer can provide (e.g., a traffic shaper may discard the traffic excess). In Release 8, the MBR of a GBR bearer shall be set equal to the GBR [1].

The last two parameters are applicable only to GBR bearers.

### 4.2 Standardized QCI characteristics

The QCI determines the packet forwarding treatment that the EPS have to apply to the traffic conveyed by the bearer. The 3GPP specification [2] defines a set of standardized QCI characteristics associated with the QCI values. Three of these characteristics are relevant for scheduling:

- *Resource type (GBR or non-GBR)*: determines if the bearers associated to a given QCI are GBR or not.
- *Priority*: via its QCI, a bearer is associated with a priority level, which can be used to prioritize between bearers. Note that a QCI with lower priority parameter has preference over a QCI with higher priority parameter.
- *Packet Delay Budget (PDB)*: it defines the upper limit of the delay suffered by a packet between the UE and the Policy and Charging Enforcement Function (PCEF).

Table 2 captures the mapping of standardized QCI values and its corresponding characteristics as well as example services for each QCI.

It is relevant to note that according to [1] these standardized characteristics are simply criteria for the configuration of parameters in each node for each QCI. However, they are not signaled on any interface. Additionally, it is also worth mentioning that the PDB limits the total delay between the PCEF and the UE, and therefore, to calculate the budget applicable to the radio access network, the delay between the PCEF and the eNodeB must be subtracted. Hereafter, we will refer to the PDB assuming that the delay between the PCEF and the eNodeB has already been subtracted.

### 4.3 Implications of the bearer QoS profile on the scheduling framework

Let us further analyze the implications of the 3GPP QoS concept on the scheduling solution.

Regarding GBR bearers, a 3GPP-QoS-compliant scheduling framework must be able to guarantee the GBR for the bearers with QCIs 1–4. As for Release 8, the MBR shall be set equal to the GBR and the bearer must be able to guarantee the bit rate up to the limit imposed by the MBR. As stated in [2], services mapped onto these QCIs “can assume that congestion related packet drops will not occur, and 98 percent of the packets shall not experience a delay exceeding the QCI’s PDB” (except in the case of transient link outages). Therefore, we conclude that, for a GBR bearer, the scheduling framework must guarantee that the packet delay does not exceed the PDB for all incoming traffic up to the limit of the GBR (for at least 98% of the packets).

Regarding non-GBR bearers, as stated above, the network has no obligation to guarantee a given bit rate to the EPS bearers. Then, it raises the question of how should the PDB attribute be interpreted for non-GBR bearers that are expected to support elastic traffic. If a communication link supports traffic (e.g., elastic TCP-based) with an upper bound on link capacity, this path is not able to guarantee a given target delay for that traffic unless a restrictive rate shaping function (e.g., leaky bucket) is applied [18]. However, in LTE, the MBR parameter is not applicable to non-GBR bearers. In this respect, the 3GPP specification [2] states that “in general, the rate of congestion related packet drops can not be controlled precisely for Non-GBR traffic.” Additionally, it indicates that a queue management function can contribute to control the packet drop rate. Regarding the satisfaction of the PDB for non-GBR traffic 3GPP, the specification [2] states that “98 percent of the packets that have not been dropped due to congestion should not experience a delay exceeding the QCI’s PDB.” Therefore, ultimately, we conclude that (i) the 3GPP specification acknowledges that non-GBR bearers (e.g.,

**Table 2** Standardized QCI values to standardized characteristic mapping [2]

QCI	Resource type	Priority	Packet Delay Budget (ms)	Packet error loss rate	Example services
1	GBR	2	100	$10^{-2}$	Conversational voice
2		4	150	$10^{-3}$	Conversational video (live streaming)
3		3	50	$10^{-3}$	Real-time gaming
4		5	300	$10^{-6}$	Non-conversational video (buffered streaming)
5		1	100	$10^{-6}$	IMS signalling Video (buffered streaming), TCP-based (e.g.,
6	Non-GBR	6	300	$10^{-6}$	www, e-mail, chat, ftp, p2p file sharing, and progressive video)
7		7	100	$10^{-3}$	Voice, video (live streaming), interactive gaming
8		8	300	$10^{-6}$	Video (buffered streaming), TCP-based (e.g.,
9		9			www, e-mail, chat, ftp, p2p file sharing, and progressive video)

supporting elastic traffic) are expected to suffer packet drops due to congestion and (ii) the delay of packets that are not dropped should not exceed the PDB, but this delay limit is not as strict as for GBR bearers.

Regarding the relative priority between bearers of different QCIs, the 3GPP specification [2] indicates that the scheduling between different bearers “shall primarily be based on the PDB.” However, if the PDB cannot be met for all bearers with sufficient radio channel quality then Priority shall be used as follows [2]: “in this case a scheduler shall meet the PDB of a bearer on Priority level  $N$  in preference to meeting the PDB of a bearer on Priority level  $N+1$ .” Therefore, we conclude that a 3GPP-based scheduling framework has to be able to detect if the PDB cannot be fulfilled for all bearers (e.g., in case of heavy load) and, in that case, prioritize between bearers of different QCIs.

Additionally, it is relevant to observe that [2] states that the prioritization should not be triggered if the PDB is not fulfilled by a UE with insufficient radio channel quality. Although [2] does not define any criterion to determine if a UE has sufficient channel quality, it is possible to use UE measurements to establish a radio channel criterion.

## 5 QoS scheduling framework

This section presents the scheduling framework based on the 3GPP guidelines and EPS bearer QoS attributes described above. First, for non-GBR bearers, this proposal includes a utility maximization scheduling discipline (based on [16]). Second, for GBR bearers, the proposal includes a delay-dependent scheduling discipline combined with a rate shaping function. Third, the proposal includes a novel algorithm that establishes relative precedence between QCIs when the PDB can no longer be met for all bearers.

### 5.1 Scheduling for non-GBR bearers

As it can be seen in Table 2, the TCP elastic traffic is to be mapped onto non-GBR bearers. The literature on scheduling for elastic traffic in OFDMA systems is extensive. Some of these works have designed scheduling disciplines aiming at maximizing the sum (over flows) of the average rate of the flows under power and/or minimum rate constraints [19–23]. Other works have concentrated on maximizing the sum of a utility function instead [24, 25]. However, it is not possible for a scheduling discipline to guarantee a target delay for elastic traffic unless a restrictive rate shaping function (e.g., leaky bucket) is applied [18].

On the other hand, the quality of TCP-based services is typically determined by the throughput of the flow. For example, for interactive or background services, the quality typically depends on the service response time, which is ultimately determined by the throughput of the flow (see, for example, [26] for web browsing). For progressive video, the quality primarily depends on the rebuffering events [27], which are also determined by the throughput of the flow [28, 29].

For these reasons, we choose to schedule the traffic mapped on non-GBR bearers with a utility maximization scheduling discipline where the utility function is the  $\alpha$ -fair function [16]. The scheduling metric is computed as follows:

$$P_k[n, s] = \frac{R_k[n, s]}{\left[\overline{r_k[n]}\right]^\alpha} \quad (3)$$

where  $P_k[n, s]$  represents the priority of bearer  $k$  on resource block  $s$  and TTI  $n$ ,  $R_k[n, s]$  denotes the achievable transmission rate (obtained from the Channel Quality Indicator),  $\overline{r_k[n]}$  is the low pass filtered data rate that bearer  $k$  has received until TTI  $n$  [24], and  $\alpha$  is a factor

that controls the degree of fairness. Results presented in [16] have shown that by setting  $\alpha \approx 0.6$  it is possible to achieve a user throughput at 5 % outage similar to the Proportional Fair algorithm but an average user throughput gain of approximately 60 %. Therefore, hereafter, it will be assumed that  $\alpha$  is set to 0.6.

## 5.2 Scheduling for GBR bearers

As described in Section 4.3, the eNodeB has to ensure for GBR bearers that 98 % of the packets shall not suffer a delay that exceeds the QCI's PDB. This requires a tight control of the delay suffered by the packets in the eNodeB queues. Interesting QoS schedulers are the M-LWDF, Exp rule, and Log rule [3]. As example, the Exp rule scheduler computes the user's priority as:

$$P_k[n, s] = b_k \cdot R_k[n, s] \cdot \exp\left(\frac{\tilde{a}_k w_k[n]}{1 + \sqrt{(1/K) \sum_k w_k[n]}}\right) \quad (4)$$

for any fixed positive  $b_k$  and  $\tilde{a}_k$ .  $w_k[n]$  represents the HOL delay of user  $k$  measured in TTI  $n$ . The parameter  $b_k$  can be set equal to  $1/r_k[n]$ . This way, the priority is computed as in the Proportional Fair discipline and then multiplied by a delay-dependent factor. Additionally, the HOL packet delay  $w_k[n]$  can be substituted by the queue length, which yields its queue-length-driven version.

While the Exp rule and Log rule scheduling algorithms provide interesting performance benefits [3–5], they are not well suited for fulfilling the QoS requirements of GBR bearers exposed in Section 4.3. Firstly, although these strategies increase the user's priority when the HOL packet delay increases, they may be enhanced by emphatically increasing the user's priority when the HOL packet delay approaches its PDB. Secondly, and more importantly, the upper bound of the delay-dependent factor in these strategies cannot be easily controlled. This upper bound is reached when  $w_k[n]$  approaches the PDB (e.g., due to a temporary load increase). For example, in (4), this upper bound depends on the PDB of bearer  $k$  and the average HOL packet delay of all bearers. This makes it difficult to eventually prioritize bearers of a given QCI at the expense of bearers mapped onto QCIs with a higher priority parameter if the latter bearers suffer a large HOL packet delay.

For these reasons, we propose to implement the delay-dependent factor based on a sigmoid function [30] instead of the exponential or logarithmic functions. The proposed delay dependent factor is as follows:

$$f(w_k) = \frac{c}{1 + e^{-a_k(w_k - D)}} \quad (5)$$

The parameter  $a_k$  adjusts the slope of the sigmoid function, the parameter  $c$  establishes its upper bound, and the parameter  $D$  can be used to control the target packet

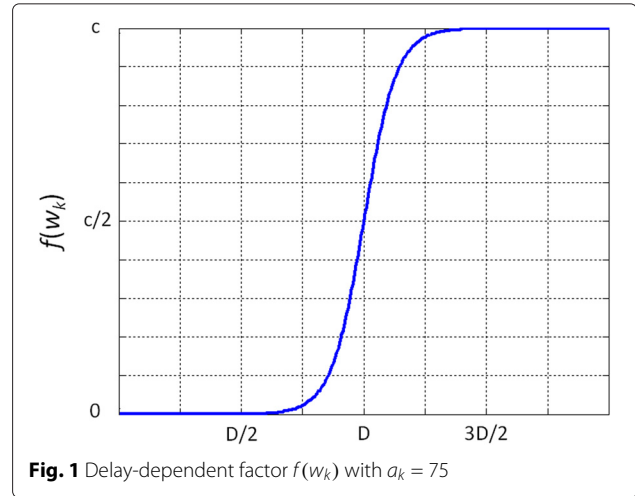


Fig. 1 Delay-dependent factor  $f(w_k)$  with  $a_k = 75$

delay. Although the sigmoid utility function is applied with respect to the bandwidth resource in [30], we apply it with respect to the packet delay to implement the delay-dependent factor. Figure 1 depicts  $f(w_k)$  for the case  $a_k = 75$ .

Our scheduling proposal for GBR bearers computes the priority of bearer  $k$  by combining the delay-dependent factor  $f(w_k)$  with the scheduling metric for non-GBR bearers of (3) as follows:

$$P_k[n, s] = (1 + f(w_k)) \cdot \frac{R_k[n, s]}{\left[r_k[n]\right]^\alpha} \quad (6)$$

$P_k[n, s]$  in (6) provides a scheduling priority similar to non-GBR bearers (see Section 5.1) when the Head of Line delay  $w_k \rightarrow 0$ , but it increases when the Head of Line delay  $w_k$  approaches  $D$ . Hence, by appropriately setting the parameters  $a_k$ ,  $D$ , and  $c$ , (6) is able to emphatically increase the priority  $P_k[n, s]$  of GBR users when their Head of Line delays approach their respective PDB. With this proposal, if the Head of Line delay  $w_k$  approaches the PDB (e.g., due to a temporary load increase), the delay-dependent factor will be upper bounded by parameter  $c$ , and therefore, it will cause a controlled increase of the priority  $P_k[n, s]$ . Furthermore, to avoid that an excessive reduction of  $r_k[n]$  caused by the starvation of a user also leads to an uncontrollable increase of  $P_k[n, s]$ , the metric  $r_k[n]$  will be lower bounded by a parameter  $r_{lb}$ .

Besides the scheduling discipline controlled by the scheduling metric (6), we propose that a rate shaping function (e.g., a leaky bucket) limits the maximum bit rate of the traffic flow as imposed by the MBR parameter. The combination of the rate shaping function and the satisfaction of the PDB guarantees the satisfaction of the GBR parameter.

### 5.3 QCI prioritization

In this subsection, we present an algorithm for integrating the above described scheduling disciplines for GBR and non-GBR bearers. Additionally, the algorithm implements the relative prioritization between QCIs described in Section 4.3.

The integration mechanism modifies the priority of a bearer  $k$  that is mapped onto  $QCI_m$  by multiplying it by a factor  $F_k^{QCI}$ . The new priority of the bearer  $k$  is given by:

$$P_k^{QCI_m}[n, s] = P_k[n, s] \cdot F_k^{QCI_m} \tag{7}$$

Then, the scheduler algorithm assigns resource block  $s$  in TTI  $n$  to the user  $k'$  with the highest priority  $P_k^{QCI_m}[n, s]$ :

$$k'[n, s] = \operatorname{argmax}_{k \in K} [P_k^{QCI_m}[n, s]] \tag{8}$$

The factor  $F_k^{QCI_m}$  implements the relative prioritization between QCIs. It is computed in every TTI and  $\forall k \in K$ . The calculation of  $F_k^{QCI_m}$  is described next.

As described in Section 4.3, the scheduling has to detect if all bearers fulfill their target quality. For this purpose, we

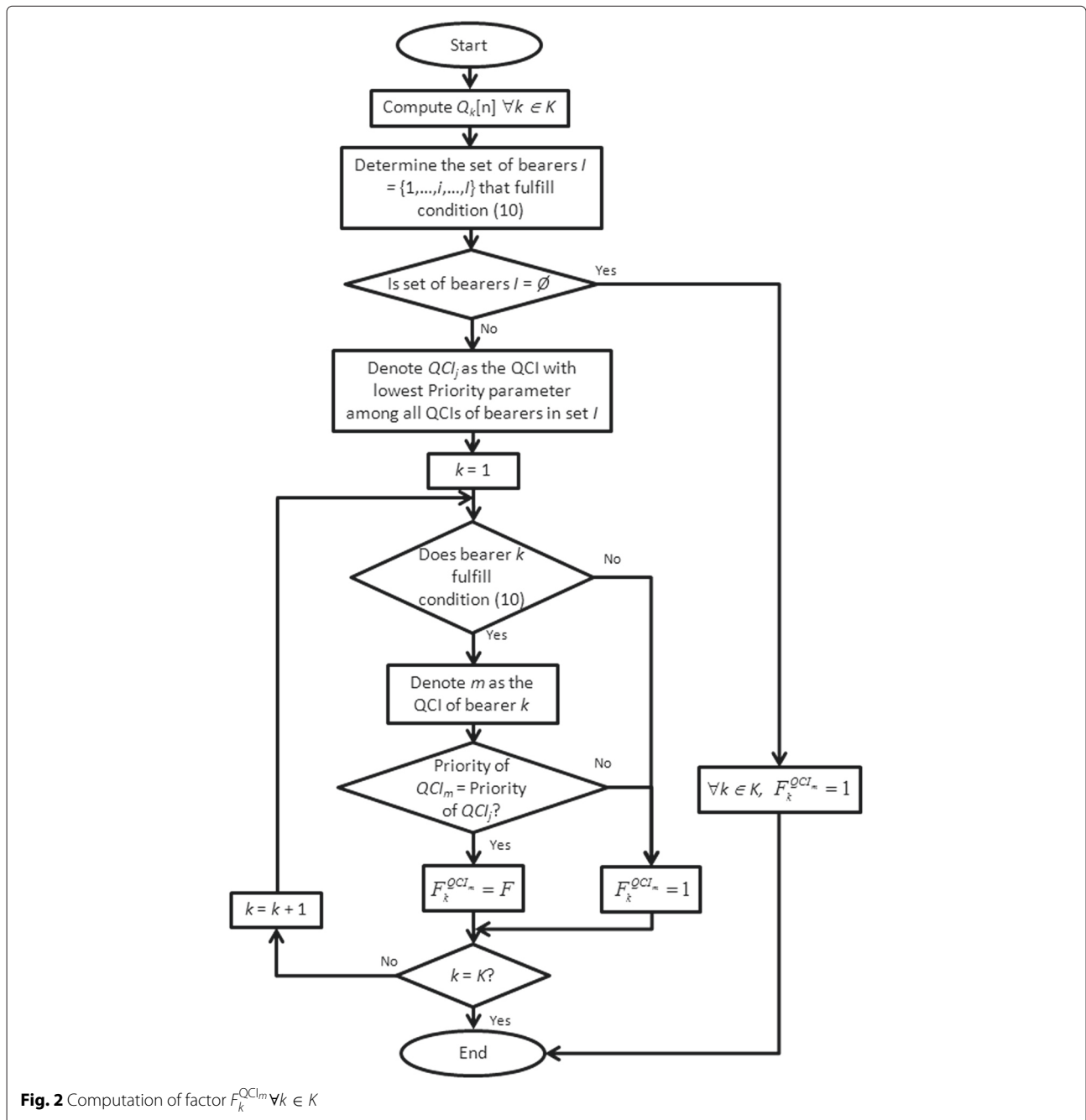


Fig. 2 Computation of factor  $F_k^{QCI_m} \forall k \in K$

define a quality performance indicator  $Q_k$  for each bearer. For bearers mapped onto GBR QCIs,  $Q_k$  is an estimator of the average packet delay of bearer  $k$ . For bearers mapped onto non-GBR QCIs,  $Q_k$  is an estimator of the average data rate of bearer  $k$ .  $Q_k$  is then defined as follows:

$$Q_k[n] = \begin{cases} \bar{d}_k[n] = (1 - \rho_d) \cdot \bar{d}_k[n-1] + \rho_d \cdot \frac{q_k[n]}{\lambda_k} \\ \text{QCI}_m = 1, 2, 3, 4 \\ \bar{r}_k[n] = (1 - \rho_r) \cdot \bar{r}_k[n-1] + \rho_r \cdot r_k[n] \\ \text{QCI}_m = 5, 6, 7, 8, 9 \end{cases} \quad (9)$$

where  $q_k[n]$  denotes the number of bits in the queue of bearer  $k$  in TTI  $n$  and  $\lambda_k$  is an estimator of the average arrival bit rate on bearer  $k$ . Accordingly,  $r_k[n]$  represents the transmitted data rate in TTI  $n$  by bearer  $k$ .  $\rho_d$  and  $\rho_r$  are time averaging constants.

Let us additionally define a target quality  $TQ^{\text{QCI}_m}$  that establishes the minimum quality level that should be experienced by a bearer mapped onto  $\text{QCI}_m$ :

- For GBR QCIs:  $TQ^{\text{QCI}_m}$  is a delay threshold for  $\text{QCI}_m$ .
- For non-GBR QCIs:  $TQ^{\text{QCI}_m}$  is a data rate threshold for  $\text{QCI}_m$ .

Then, we define condition (10) that is fulfilled if bearer  $k$  satisfies

$$\begin{cases} TQ^{\text{QCI}_m} < Q_k[n] & \text{QCI}_m = 1, 2, 3, 4 \\ Q_k[n] < TQ^{\text{QCI}_m} & \text{QCI}_m = 5, 6, 7, 8, 9 \end{cases} \quad (10)$$

A bearer  $k$  fulfills condition (10) if its quality performance indicator  $Q_k[n]$  does not reach the target quality level. If there exists a bearer  $k \in K$  that fulfills condition (10), then CQI prioritization must be triggered.

Figure 2 depicts the algorithm proposed to compute the factor  $F_k^{\text{QCI}_m} \forall k \in K$ . The algorithm is based on condition (10), and it has the following steps:

1. The algorithm determines the set  $I = \{1, \dots, i, \dots, I\}$  of bearers that fulfill condition (10).
2. If set  $I$  is empty, CQI prioritization is not required. Then,  $F_k^{\text{QCI}_m} = 1 \forall k$ , and the algorithm is finished. If set  $I$  is not empty, CQI prioritization is triggered. Then, the algorithm starts to compute factor  $F_k^{\text{QCI}_m}$  for  $k = 1$ .
3. The algorithm sets factor  $F_k^{\text{QCI}_m}$  for bearer  $k$  mapped on to  $\text{QCI}_m$  equal to a constant factor  $F$  if the following two conditions hold: (i) bearer  $k$  fulfills condition (10) and (ii)  $\text{QCI}_m$  is the QCI with the lowest priority parameter among the QCIs of all the bearers  $\in I$ .
4. The algorithm repeats step 3 until  $k$  is the last bearer  $\in K$ .

In summary, factor  $F_k^{\text{QCI}_m}$  is set equal to a priority enhancing factor  $F$  if bearer  $k$  fulfills condition (10) and its QCI is the one with the lowest priority parameter among all bearers that fulfill (10). Bearers with insufficient radio channel quality are not considered in the prioritization algorithm and are assumed to set  $F_k^{\text{QCI}_m} = 1$ .

We further propose to provide a certain guard margin to this reactive mechanism before the quality target limit is exceeded. For example, for QCIs = 1, 2, 3, 4, if the PDB of a given QCI equals 300 ms, then we set  $TQ^{\text{QCI}_m}$  to a lower value (e.g., 250 ms), thereby providing a certain operation margin (50 ms) before the PDB is exceeded. See Fig. 3 for the relation between parameters  $D$ ,  $TQ^{\text{QCI}_m}$ , and PDB. Accordingly, for QCIs = 5, 6, 7, 8, 9, if UEs are considered to incur outage at a given throughput threshold (e.g., 512 kbps), we establish  $TQ^{\text{QCI}_m}$  (e.g., 600 kbps) including a certain guard interval.

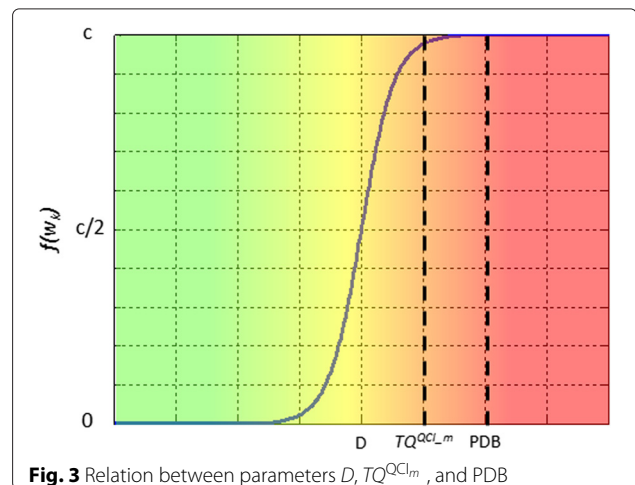
The complexity of the overall scheduling solution is similar to that of the Exp rule scheduler. It has a complexity of  $O(KS)$  due to a search for the maximum of  $K$  metrics on each of the  $S$  resource blocks. The CQI Prioritization algorithm has a low complexity as it only requires the evaluation of condition (10) for each of the  $K$  bearers and determining which bearers in set  $I$  have the lowest priority parameter.

## 6 Performance results

This section presents the performance evaluation of the proposed scheduling framework in a quasi-dynamic system level simulator. We compare the results with the Proportional Fair and Exp rule scheduling algorithms.

### 6.1 Simulation setup and parameters

We consider a hexagonal network of 13 cells. We simulate users only in the central cell of the grid, whereas the remaining cells are a source of interference. Users maintain their geographical location during their lifetime, and



**Fig. 3** Relation between parameters  $D$ ,  $TQ^{\text{QCI}_m}$ , and PDB



consequently, their deterministic path loss and shadow fading do not vary. However, users suffer a fast fading process that is updated in each TTI. We use the ITU *Typical Urban* (TU) power delay profile as the multi-path model. The Geometry Factor distribution provided by the simulation model precisely matches the results for a macro-cell outdoor scenario presented in [31]. We use the Exponential Effective SIR Metric (EESM) to model the link-to-system-level mapping.

Four different services are considered: live video streaming, buffered video streaming, web browsing, and ftp. These services are mapped onto QCIs 2, 4, 8, and 9, respectively (see Table 2). The user's birth process follows a Poisson process. We control the mean offered cell load with a simulation parameter. Unless otherwise stated, the mean offered cell load is set to 11.5 Mbps, which is a heavy load setting for the considered network

configuration. Live video streaming and buffered video streaming services are modeled as CBR sources with bit rates 240 and 440 kbps, respectively. We consider a simplified web browsing and ftp models: each user downloads just one web page (of size 250 kB) or ftp file (of size 1 MB) in every session, and when the download is completed, the session is finished. Five different cases of traffic mix shares are considered (see Table 3). Users with G Factor  $\leq -3$  dB are assumed to have insufficient radio channel quality, and therefore, they are not considered in the prioritization algorithm (see Section 5.3). The remaining parameters of the simulated LTE model are included in Table 3.

## 6.2 Simulation results

### 6.2.1 Sensitivity to prioritization factor $F$

We commence the results section by analyzing the sensitivity of the prioritization mechanism between QCIs

**Table 3** 3G LTE network model

Parameter	Setting
Carrier frequency	2 GHz
System bandwidth	10 MHz
Inter-site distance	2 km
Std of shadow fading	8 dB
Power delay profile	ITU typical urban 20 paths
Antenna scheme	SIMO 1 × 2 – rx (MRC)
UE speed for fast fading process	3 km/h
Total eNode-B transmit power	46 dBm
Pilot, control channel overhead	3/7 symbols
Modulation/coding rate settings	QPSK: 1/3, 1/2, 2/3 16QAM: 1/2, 2/3, 4/5 64QAM: 2/3, 4/5
CQI estimation	Ideal (std of estimation error 0 dB)
HARQ model	Ideal chase combining
Scheduling parameters	$\alpha_k = 75, c = 4, 3D/2 = \text{PDB}$ $r_{lb} = 100 \text{ kbps}, \rho_d = 1/100$ $\rho_r = 1/300, F = 4 \text{ dB}$
Insufficient radio channel quality	G Factor $\leq -3$ dB
Services	QCI <sub>2</sub> : live video streaming QCI <sub>4</sub> : buffered video streaming QCI <sub>8</sub> : web QCI <sub>9</sub> : ftp TQ <sup>QCI<sub>2</sub></sup> = 125 ms, PDB = 150 ms TQ <sup>QCI<sub>4</sub></sup> = 250 ms, PDB = 300 ms TQ <sup>QCI<sub>8</sub></sup> = 600 kbps TQ <sup>QCI<sub>9</sub></sup> = 600 kbps
Traffic Mix Cases	Equal Load 25 % load per service Heavy Load live video streaming 40 % live video 20 % other services Heavy Load buffered video streaming 40 % buffered video 20 % other services Heavy Load web 40 % web, 20 % other services Heavy Load ftp 40 % ftp, 20 % other services

to the factor  $F$ . As described in Section 5.3, factor  $F_k^{QCI_m}$  is set equal to the priority enhancing factor  $F$  if bearer  $k$  fulfills condition (10) and its QCI is the one with the lowest priority parameter among all bearers that fulfill (10).

Figure 4 depicts key performance indicators for the proposed QoS scheduler and for the traffic mix case Equal Load. For QCI<sub>2</sub> and QCI<sub>4</sub>, the figure represents the percentage of user entities (UEs) with sufficient channel quality for which at least 2% of the packets suffer a delay exceeding the QCI's PDB. For QCI<sub>8</sub> and QCI<sub>9</sub>, Fig. 4 represents the percentage of UEs with sufficient channel quality that have a throughput below 512 kbps.

When the factor  $F$  increases, the relative prioritization between QCIs is more aggressive. Therefore, the percentage of users of QCI<sub>9</sub> in outage increases (see Fig. 4). On the other hand, a low factor  $F$  is unable to prioritize QCIs. This can be observed for QCI<sub>8</sub>, in which the percentage of users that suffer a throughput below 512 kbps increases when  $F$  decreases below 3–4 dB. Additionally, when  $F$  is higher than 4–5 dB, the percentage of users of QCI<sub>2</sub> and QCI<sub>4</sub> that incur outage increases. A very high  $F$  penalizes the efficiency of the scheduling algorithm to serve users in the fading tops. For these reasons, we select a factor  $F = 3$ –4 dB for the design.

### 6.2.2 Evaluation with different traffic mixes

Table 4 compares key performance indicators between the Proportional Fair, the Exp rule scheduler, and the proposed QoS scheduler for all considered traffic mix cases (H.L. stands for Heavy Load). The implementation of the Exp scheduler assumes that the priority of GBR flows is computed using (4), whereas the priority of non-GBR flows is computed using Proportional Fair. The parameters used in (4) are extracted from [3].

The results show that more than 99% of the users of QCI<sub>2</sub> fulfill the PDB for the considered traffic mix cases and for the three schedulers due to the low input rate of the considered live streaming service. However, for QCI<sub>4</sub> the percentage of users that do not fulfill the PDB varies in the different traffic mix cases from approximately 4.5 to 7.1% with Proportional Fair. With the Exp rule and the QoS scheduler, this outage is almost eliminated.

Regarding QCI<sub>8</sub>, the proposed QoS scheduler considerably reduces the percentage of users that suffer a throughput below 512 kbps compared to both the Proportional Fair and especially the Exp rule scheduler. On the other hand, for QCI<sub>9</sub>, the proposed QoS scheduler increases the percentage of users in outage for all traffic mix cases. As QCI<sub>9</sub> has the highest priority parameter, when the target

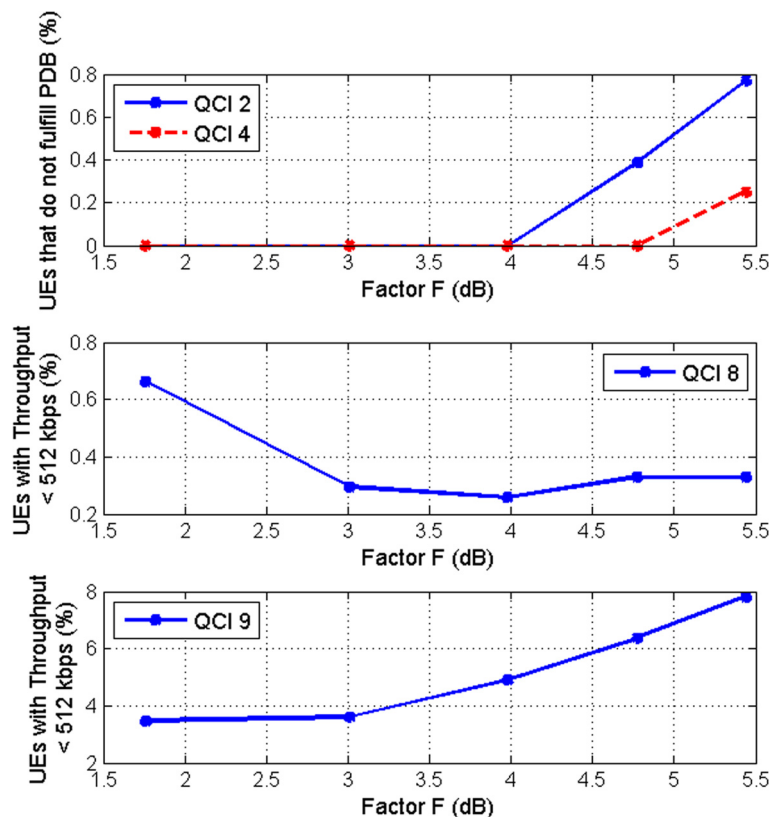


Fig. 4 Performance sensitivity to prioritization factor  $F$

**Table 4** Performance indicators of simulation results

			QCI <sub>2</sub>		QCI <sub>4</sub>		QCI <sub>8</sub>			QCI <sub>9</sub>		
Cell throughput (Mbps)			UEs that do not fulfill PDB (%)	Avg. UE throughput	UEs that do not fulfill PDB (%)	Avg. UE throughput (kbps)	UEs with throughput < 512 kbps (%)	Avg. UE throughput (Mbps)	UE throughput at 95th percentile (Mbps)	UEs with throughput < 512 kbps (%)	Avg. UE throughput (Mbps)	UE throughput at 95th percentile (Mbps)
Equal Load	PF	11.48	0.85	240	4.79	437	2.09	3.18	10.59	2.89	2.63	7.81
	QoS	11.48	0	239	0	434	0.25	4.04	16.02	4.92	3.12	10.26
	Exp	11.48	0	241	0	437	4.67	2.97	10.20	4.49	2.47	7.56
H. L. live video	PF	11.60	0.08	240	5.16	439	1.99	3.01	9.58	2.20	2.47	5.86
	QoS	11.60	0	240	0	436	0.40	3.91	14.63	4.4	3.00	9.09
	Exp	11.60	0	240	0	439	6.16	2.71	8.88	6.05	2.25	5.60
H. L. buff. video	PF	11.77	0	240	7.14	438	3.74	2.92	9.80	4.11	2.12	5.80
	QoS	11.77	0.48	239	0.58	435	3.34	3.60	14.83	11.98	2.40	7.74
	Exp	11.77	0	240	0	439	10.44	2.61	9.31	12.70	1.91	5.23
H. L. web	PF	11.27	0	240	4.57	439	1.66	3.57	11.70	1.65	2.97	8.14
	QoS	11.27	0	239	0	439	0.67	4.56	17.41	5.69	3.56	11.45
	Exp	11.27	0	240	0	440	3.75	3.41	11.36	3.49	2.85	8.15
H. L. ftp	PF	11.67	0.49	240	6.09	438	2.61	3.30	10.86	3.04	2.78	8.34
	QoS	11.67	0	239	0	434	0	4.09	16.14	6.44	3.20	10.9
	Exp	11.67	0	240	0	438	5.54	3.10	10.32	6.18	2.66	7.80

quality  $TQ^{QCI_m}$  cannot be fulfilled for all users, the factor  $F_k^{QCI_m}$  of the QoS scheduler prioritizes users mapped onto  $QCI_2$ ,  $QCI_4$ , and  $QCI_8$ . This occurs at the expenses of users mapped onto  $QCI_9$ .

Additionally, Table 4 shows that the proposed QoS scheduler increases the average and the 95th percentile of the user throughput for  $QCI_8$  and  $QCI_9$ .

Let us define a metric  $Q^{QCI_m}[n]$  that measures the performance of the user with the worst quality among all users mapped on to  $QCI_m$  that have sufficient radio channel quality:

$$Q^{QCI_m}[n] = \max_{\forall k \in QCI_m} Q_k[n] \quad QCI_m = 1, 2, 3, 4. \quad (11)$$

Figure 5 plots a realization of the evolution of  $Q^{QCI_m}[n]$  for  $QCI_4$  in the Equal Load case for the Proportional Fair and the proposed QoS scheduling algorithms. For  $QCI_4$ ,  $Q^{QCI_m}$  measures the filtered packet delay of the bearer with minimum  $Q_k[n]$  quality among all bearers of  $QCI_m$ . It can be observed how the QoS scheduling algorithm is able to keep the filtered packet delay of all users of  $QCI_4$  below their PDB.

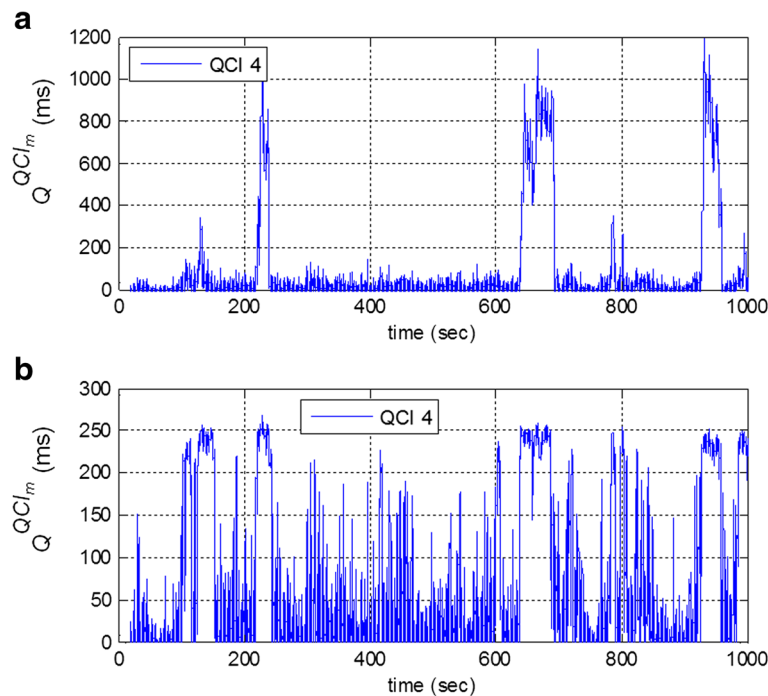
### 6.2.3 Evaluation under different offered loads

Here we analyze the performance of the proposed scheduling framework for different offered loads. For the evaluation, we consider the traffic mix case Equal Load. For  $QCI_2$  and  $QCI_4$ , Fig. 6 depicts the percentage of users with sufficient channel quality for which at least 2 % of the

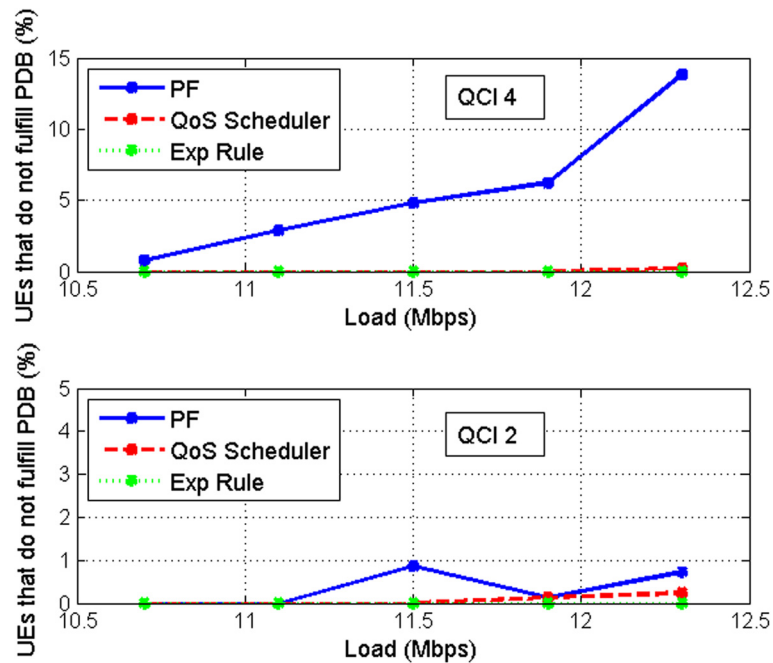
packets suffer a delay exceeding the QCI's PDB. For  $QCI_8$  and  $QCI_9$ , Fig. 7 represents the percentage of users with sufficient channel quality that suffer a throughput below 512 kbps. The results show that under a wide range of load conditions the proposed scheduler is able to eliminate or significantly reduce the outages of  $QCI_2$ ,  $QCI_4$ , and  $QCI_8$ . This occurs at the expenses of increasing the outage of  $QCI_9$ . The results also show that the Exp rule is also robust with the outage of GBR users, but it degrades considerably the performance of users in  $QCI_8$  and  $QCI_9$ . This is not in line with 3GPP QoS that states that users of  $QCI_8$  should be prioritized over users of  $QCI_9$ .

## 7 Conclusions

In this paper, we present a scheduling framework with the objective of globally addressing the QoS requirements as defined by 3GPP specifications. For this, we have analyzed the implications of the 3GPP QoS architecture on the downlink scheduling at the eNodeBs of an LTE network. For GBR bearers, the scheduler has to guarantee a packet delay below the QCI's PDB for all incoming traffic up to the limit of the GBR parameter. For non-GBR bearers, 3GPP specifications admit that congestion-related packet drops may occur and only the packets that have not been dropped due to congestion should not experience a delay exceeding the QCI's PDB, although this delay limit is not as strict as for GBR bearers. Additionally, if the PDB cannot be fulfilled for all bearers, the scheduler should prioritize between bearers of different QCIs.



**Fig. 5** Time evolution of maximum packet delay ( $Q^{QCI_m}$ ) for  $QCI_4$ . **a** Proportional Fair and **b** QoS scheduler

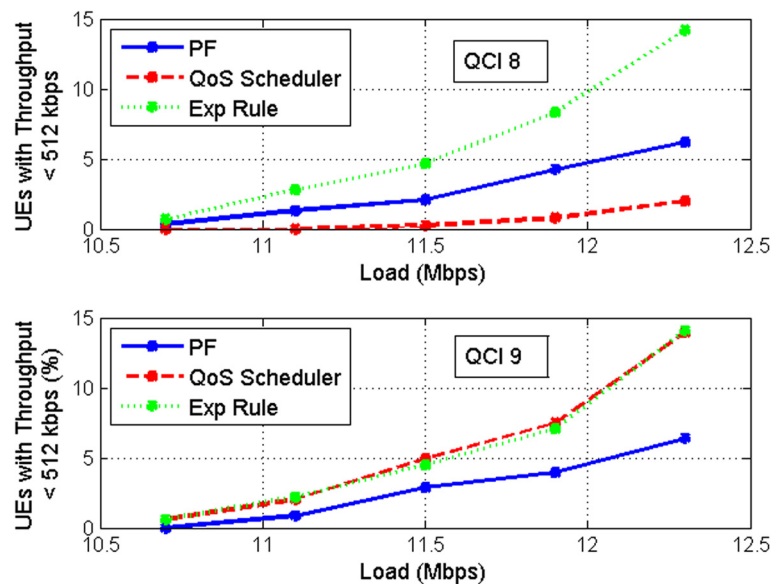


**Fig. 6** Percentage of UEs for which at least 2% of the packets suffer a delay exceeding the QCI's PDB. QCI<sub>2</sub> and QCI<sub>4</sub>

Based on the analysis of the 3GPP specifications, we have proposed a scheduling design to address the aforementioned QoS requirements. For GBR bearers, we have proposed a scheduling discipline that incorporates a delay-dependent factor based on a sigmoid function that emphatically increases the bearers' priority when the Head of Line delay approaches the PDB and that can be combined with the relative prioritization between QCIs. For non-GBR bearers, we have proposed to use a classical

channel-aware scheduling policy that aims at maximizing the sum of the concave ( $\alpha$ -fair) utility function. Additionally, we have designed a mechanism that integrates the scheduling disciplines and triggers the relative prioritization between QCIs when all bearers do not fulfill their target quality.

The simulation results have shown how the proposed scheduler is able to fulfill the PDB for GBR bearers and provides a spectrally efficient performance for non-GBR



**Fig. 7** Percentage of UEs with throughput below 512 kbps. QCI<sub>8</sub> and QCI<sub>9</sub>

bearers (average user throughput gain in the range 15–25 % over Proportional Fair). The results have also shown the ability of the proposed scheduler to prioritize bearers according to the QCI's priority if the target quality cannot be met for all bearers.

As future work, we propose a theoretical evaluation of the proposed solution.

#### Competing interests

The authors declare that they have applied for a patent relating to the content of the manuscript.

#### Acknowledgements

This work is partially supported by the Spanish Ministry of Economy and Competitiveness and the European Regional Development Fund (project TIN2013-46223-P). Additionally, the authors would like to express their gratitude to the Supercomputing and Bioinnovation Center (SCBI) of the University of Malaga (Spain) for their support and resources.

#### Author details

<sup>1</sup>Department of Signal Theory, Telematics and Communications, University of Granada, Periodista Daniel Saucedo Aranda s/n, 18071 Granada, Spain.

<sup>2</sup>Telefonica R&D - GCTO, C/ Zurbarán 12, 28010 Madrid, (Spain). <sup>3</sup>Telefonica R&D - GCTO, Ronda de la Comunicación s/n, 28050 Madrid, (Spain).

Received: 15 August 2015 Accepted: 24 February 2016

Published online: 10 March 2016

#### References

1. TS 23.401 V8.12.0. 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; General Packet Radio Service (GPRS) Enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access (Release 8) (2010). [http://www.3gpp.org/ftp/Specs/archive/23\\_series/23.401/23401-8c0.zip](http://www.3gpp.org/ftp/Specs/archive/23_series/23.401/23401-8c0.zip)
2. TS 23.203 V8.11.0. 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Policy and Charging Control architecture (Rel. 8) (2010). [http://www.3gpp.org/ftp/Specs/archive/23\\_series/23.203/23203-8b0.zip](http://www.3gpp.org/ftp/Specs/archive/23_series/23.203/23203-8b0.zip)
3. B Sadiq, R Madan, A Sampath, Downlink scheduling for multiclass traffic in LTE. *EURASIP Journal on Wireless Communications and Networking*. **2009**, 1–18 (2009)
4. F Capozzi, G Piro, LA Grieco, G Boggia, P Camarda, Downlink packet scheduling in LTE cellular networks: key design issues and a survey. *Commun. Surv. Tutor. IEEE*. **15**(2), 678–700 (2013)
5. S Dardouri, R Bouallegue, Comparative study of downlink packet scheduling for LTE networks. *Wireless Personal Commun.* **82**(3), 1405–1418 (2015)
6. P Ameigeiras, J Wigard, P Mogensen, in *Vehicular Technology Conference, 2004. VTC2004-Fall. 2004 IEEE 60th*. Performance of the M-LWDF scheduling algorithm for streaming services in HSDPA, vol. 2, (2004), pp. 999–10032. doi:10.1109/VETECF.2004.1400171
7. K Sandrasegaran, HAM Ramli, R Basukala, in *Wireless Communications and Networking Conference (WCNC), 2010 IEEE*. Delay-prioritized scheduling (DPS) for real time traffic in 3GPP LTE system, (2010), pp. 1–6
8. X Wu, X Han, X Lin, in *Communications (ICC), 2015 IEEE International Conference On*. QoS oriented heterogeneous traffic scheduling in LTE downlink, (2015), pp. 3088–3093
9. Q Ai, P Wang, F Liu, Y Wang, F Yang, J Xu, *QoS-guaranteed cross-layer resource allocation algorithm for multiclass services in downlink LTE system*. (Wireless Communications and Signal Processing (WCSP), 2010 International Conference on, Suzhou, 2010), pp. 1–4. doi:10.1109/WCSP.2010.5633846
10. M Iturralde, T Ali Yahya, A Wei, A-L Beylot, in *Vehicular Technology Conference (VTC Fall), 2011 IEEE*. Performance study of multimedia services using virtual token mechanism for resource allocation in LTE networks, (2011), pp. 1–5
11. MM Nasralla, MG Martini, in *Personal Indoor and Mobile Radio Communications (PIMRC), 2013 IEEE 24th International Symposium On*. A downlink scheduling approach for balancing QoS in LTE wireless networks, (2013), pp. 1571–1575
12. G Piro, LA Grieco, G Boggia, R Fortuna, P Camarda, Two-level downlink scheduling for real-time multimedia services in LTE networks. *Multimedia IEEE Trans.* **13**(5), 1052–1065 (2011)
13. G Mongha, K Pedersen, I Kovacs, P Mogensen, in *Proceedings of the IEEE Vehicular Technology Conference, VTC Spring 2008*. QoS oriented time and frequency domain packet schedulers for the UTRAN long term evolution, (2008)
14. Y Zaki, T Weerawardane, C Gorg, A Timm-Giel, in *Proceedings of the IEEE Vehicular Technology Conference, VTC Spring 2011*. Multi-QoS-aware fair scheduling for LTE, (2011)
15. J Góra, QoS-aware resource management for LTE-Advanced relay-enhanced network. *EURASIP J. Wireless Commun. Netw.* **2014**, 178 (2014)
16. P Ameigeiras, Y Wang, J Navarro-Ortiz, P Mogensen, JM Lopez-Soler, Traffic models impact on OFDMA systems design. *EURASIP J. Wireless Commun. Netw.* **2012**, 61 (2012)
17. M Olson, S Sultana, S Rommer, L Frid, C Mulligan, *SAE and the evolved packet core*. (Elsevier, Burlington, 2009)
18. NR Figueira, J Pasquale, An upper bound delay for the virtual-clock service discipline. *IEEE/ACM Trans. Netw.* **3**(4), 399–408 (1995)
19. J Jang, KB Lee, Transmit power adaptation for multiuser OFDM systems. *IEEE J. Selected Areas Commun.* **21**(2), 71–178 (2003)
20. LMC Hoo, B Halder, J Tellado, JM Cioffi, Multiuser transmit optimization for multicarrier broadcast channels: asymptotic FDMA capacity region and algorithms. *IEEE Trans. Commun.* **52**(6), 922–930 (2004)
21. YJ Zhang, KB Letaief, in *IEEE Transactions on Wireless Communications*. Multiuser adaptive subcarrier-and-bit allocation with adaptive cell selection for OFDM systems, vol. 3, (2004), pp. 1566–1575. doi:10.1109/TWC.2004.833501
22. H Yin, H Liu, An efficient multiuser loading algorithm for OFDM-based broadband wireless systems. *Proceedings of the IEEE Global Communications Conference 2000 (GLOBECOM'00)*. **1**, 103–107 (2000)
23. K Seong, M Mohseni, JM Cioffi, in *Proceedings of the IEEE International Symposium on Information Theory (ISIT'06)*. Optimal resource allocation for OFDMA downlink systems, (2006), pp. 1394–1398
24. G Song, Y Li, Utility-based resource allocation and scheduling in OFDM-based wireless broadband networks. *IEEE Commun. Mag.* **43**(12), 127–134 (2005)
25. P Svedman, SK Wilson, LJ Cimini, B Ottersten, Opportunistic beamforming and scheduling for OFDMA systems. *IEEE Trans. Commun.* **55**(5), 941–952 (2007)
26. P Ameigeiras, JJ Ramos-Munoz, J Navarro-Ortiz, P Mogensen, JM Lopez-Soler, QoE oriented cross-layer design of a resource allocation algorithm in beyond 3G systems. *Comput. Commun. J.* **33**, 571–582 (2010)
27. RKP Mok, EWW Chan, RKC Chang, in *IFIP/IEEE International Symposium on Integrated Network Management (IM)*. Measuring the quality of experience of HTTP video streaming, (2011), pp. 485–492
28. P Ameigeiras, JJ Ramos-Munoz, J Navarro-Ortiz, JM Lopez-Soler, Analysis and modeling of YouTube traffic. *Transactions on Emerging Telecommunications Technologies*, John Wiley & Sons. **23**(4), 360–377 (2012)
29. JJ Ramos-Munoz, J Prados-Garzon, P Ameigeiras, J Navarro-Ortiz, JM Lopez-Soler, Characteristics of mobile youtube traffic. *IEEE Wireless Commun. Mag.* **21**(1), 18–25 (2014)
30. W Kuo, W Liao, Utility-based radio resource allocation for QoS traffic in wireless networks. *IEEE Trans. Wireless Commun.* **7**(7), 2714–2722 (2008)
31. NW, et al., in *Proceedings of the IEEE Vehicular Technology Conference, VTC 2006-Fal*. Baseline E-UTRA Downlink spectral efficiency evaluation, (2006), pp. 1–5